ARTICLE IN PRESS

Animal Behaviour xxx (2016) e1e4

Contents lists available at ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav

Forum

Response to Kroodsma's critique of banded wren song performance research

S. L. Vehrencamp[a, b, *], S. R. de Kort[b, c], A. E. Illes[b, d]

[a] Department of Neurobiology and Behavior, Cornell University, Ithaca, NY, U.S.A.
[b] Laboratory of Ornithology, Cornell University, Ithaca, NY, U.S.A.
[c] Conservation, Evolution and Behaviour Research Group, School of Science and the Environment, Manchester Metropolitan University, Manchester, U.K.
[d] Department of Biology, Coe College, Cedar Rapids, IA, U.S.A.

a r t i c l e i n f o

Article history:
Received 28 July 2016
Initial acceptance 23 September 2016
Final acceptance 1 November 2016
Available online xxx
MS. number: AF-16-00673

Published file has been converted from pdf to Word. I offer some responses (yellow hi-lite) to red-letter thoughts .

A critical review of a popular scientific theory, large or small, is something we applaud because, if well executed, it stimulates discussion and progresses science. However, such a review needs to be balanced, objective, informed and logical, especially if it concludes that a well-supported theory is flawed. Unfortunately, Kroodsma's current criticism (Kroodsma, 2017) of the birdsong performance literature suffers from the same weaknesses as his earlier criticism of song repertoire use in sexual selection (Byers & Kroodsma, 2009), despite the fact that he has been alerted to those mistakes (Collins, de Kort, Perez-Tris, & Telleria, 2011).

I am very familiar with the Collins et al. paper. See next comments in yellow:

Those weaknesses include outright errors and misrepresentations, highly selective citation of the literature and convoluted logic (sensu Podos, 2017). Here we would like to take this opportunity to redress the specific issues he raises with respect to our work on the banded wren, Thryophilus pleurostictus, and by doing so, illustrate how his criticism is flawed as a result of the above weaknesses, his restricted definition of ‘song performance’, and a misunderstanding of the song system of the banded wren.

How an author evaluates and perceives the published literature offers insight into the filters that author applies to what is published by others. Those same perceptions also reveal the level of rigor that an author will apply to his or her own work, giving insight as to what an author herself (in this case) publishes as science. Here is no doubt the source of differences that Vehrencamp and I will have: While I find zero-support for the critiqued “theory,” Vehrencamp et al. accept it as popular and well-supported. My abundant criticisms of this work are in my Forum article and also in a more detailed response to Podos on this website.

I want to emphasize that we are not getting off to a good start. In my opinion, anyone who wholeheartedly defends this performance/deviation literature, flawed as it is, without having a single critical thought against it, is in trouble from the start. Anyone who proclaims the merits of this literature without a reasoned reflection on what is good or bad is not likely to retain a lot of credibility on the important issues. Yes, I understand that the focus of the Vehrencamp et al. rebuttal is on banded wrens, but this introductory paragraph about the general nature of the performance literature says a lot about what to expect of their own work as well. If the bar is low for accepting the conclusions of published work in the deviation/performance literature, and the barriers high for criticism of such work, then I don’t think we are going to have a useful exchange.

It’s abundantly clear that the authors feel that my view of the world is flawed and full of mistakes/weaknesses/errors/misrepresentations/convoluted logic, selective citation(s), etc.

Banded wren males possess song repertoires of approximately 25 distinct song types, which are largely shared with other males in their neighbourhood (Molles & Vehrencamp, 1999). The terminal trills of these song types vary in their trill note rate, frequency bandwidth and vocal deviation (maximal observed trill rates and maximal observed frequency bandwidths are inversely related in many songbirds, defining a negatively sloped upper limit line on a trill rate versus frequency bandwidth plot; the perpendicular

* Correspondence: S. L. Vehrencamp, Neurobiology and Behavior, Mudd Hall, Cornell University, Ithaca, NY 14853, U.S.A.
E-mail address: slv8@cornell.edu (S. L. Vehrencamp).

Please cite this article in press as: Vehrencamp, S. L., et al., Response to Kroodsma's critique of banded wren song performance research, Animal Behaviour (2016), http://dx.doi.org/10.1016/j.anbehav.2016.12.004

distance of a given trill from this line is its vocal deviation). Moreover, the trill notes themselves vary greatly in shape and complexity. Most of our research on this species has focused on the use of these song types in male–male territorial interactions (Burt & Vehrencamp, 2005; Hall, Illes, & Vehrencamp, 2006; Molles, 2006; Molles & Vehrencamp, 1999, 2001; Trillo & Vehrencamp, 2005; Vehrencamp, 2001; Vehrencamp, Ellis, Cropp, & Koltz, 2014; Vehrencamp, Hall, Bohman, Depeine, & Dalziell, 2007). We have shown that males negotiate their territorial boundaries primarily by varying short-term song type diversity and switching rate to indicate their propensity to approach, stand their ground or retreat from a territorial rival. Males also frequently song type-match each other during aggressive encounters. This primary role of song type choice does not rule out the possibility that subtle details of song structure also play a role and provide additional types of information about the sender, for both male and female receivers. The type-matching behaviour of countersinging males provides ample opportunities for receivers to compare their performance on the same song type, as proposed by Logue and Forstmeier (2008) for repertoire species.

The presence of signal component trade-offs (where two components of a signal are negatively correlated such that extreme values of one tends to inhibit extreme values of the other) sets up the potential for receivers to exert selective pressure on combinations that reveal useful information about the sender. This idea has been around for over two decades (Bradbury & Vehrencamp, 2011; Hebets & Papaj, 2005; Podos, 1997, 2017; Wells & Taigen, 1986). Whenever one observes a negative correlation between two signal components, it is worth testing this trade-off hypothesis (Podos's hypothesis 1) by looking to see whether receivers pay attention to alternative combinations of those components, and if so, whether individual variation in these combinations is associated with sender condition, context or reproductive success (Podos's hypothesis 2). The note structures of many birdsongs are obvious candidates for testing these hypotheses, because they are highly precise vocal utterances that have evolved under selective pressure from receiver responses in the contexts of territory defence and mate attraction (Collins, 2004; ten Cate, 2004). We examined

http://dx.doi.org/10.1016/j.anbehav.2016.12.004

several aspects of song performance in the banded wren, not limited to vocal deviation of trills as Kroodsma has restricted his critique to, but also the individual components of vocal deviation – trill rate and frequency bandwidth – along with trill note consistency, in multiple experimental and correlational studies. We have obtained consistent evidence that performance does matter in a repertoire species such as the banded wren.

OK, I now I need to put those red-letter words in the perspective of my world. I certainly agree that there are ideas worth testing. But given how Vehrencamp et al. view that the performance theory (deviation and all that) is "well-supported," and I have the polar opposite opinion, and given how they have obtained "consistent evidence" in their studies, and how all studies have found this kind of "consistency," I now inevitably must question every study that they have done. I am not going to actually study other papers (I'm really tired of all this), but for me my evaluation of Illes, Hall, & Vehrencamp serves as a proxy for the rest. Whether my n = 1 evaluation is fair or not, I do not know, but my gut feeling is that the level of rigor in one study (e.g., Illes et al.) will be indicative of the level of rigor in other studies by the same authors, especially by the senior investigator (Vehrencamp) who is presumably guiding the work. The sense that I can use this one paper is reinforced when the authors vigorously defend what I feel has some serious problems.

Our first indication that performance components affected male responses was obtained by Illes, Hall, and Vehrencamp (2006), building on the Ballentine, Hyman, and Nowicki (2004) study, in which songs modified to have faster (higher performance) or slower (lower performance) trill rates were simultaneously presented to territorial males. We found that most subjects initially approached the fast stimulus, but if they were exposed to a broader frequency bandwidth (lower vocal deviation or higher performance) trill, they subsequently spent less time close to the fast speaker. Kroodsma disparages the design, execution, analysis, results and conclusions of this study, and chides the many researchers who have cited the paper. We show below that each of his criticisms is either incorrect or misinformed.

Yes, I suppose I have been rather brutal. Let's see where I went wrong.

First, Kroodsma states that the playback should have been conducted with blind observers. The experimental design consisted of a two speaker set-up each broadcasting a separate stimulus. The observers were not informed about which speaker broadcasted which stimulus. Nevertheless, as Kroodsma acknowledges for the Cramer and Price (2007) study, an acute observer might discern which was which by listening, and we would have had to deafen the observers to exclude this possibility.

Is there anywhere in the methods section or anywhere else in the paper where the reader is informed that the observers were not informed about which songs were which? I don't think so. And even if the observers hadn't been told, it could be said thta they would have known anyway, given that they were listening to the songs. Telling me that it is difficult/impossible to judge the reactions of the birds blindly does not erase all of the problems that arise from judgements that are not done blindly. It's that simple. There's an extensive literature on the problems of nonblind judgments in work like this. I don't need to tell the authors that. And I don't need to be told that I've done playbacks in which I've not been blind; I acknowledge that. In the end, nonblind observations are tough to defend (but this issue is not the biggest problem with Illes et al., I feel, so best to move on at this point).

Second, we are surprised that an experienced ornithologist expresses doubt about the possibility of tracking movements of birds in their tropical deciduous forest habitat (i.e. when he writes, 'Even though flagging is used to mark area boundaries, the task of monitoring the location of a moving bird in this habitat seems a high challenge to accomplish with much confidence'; Kroodsma, 2017, page e13). We always had three observers for these trials, and they were all well-trained, experienced field assistants with keen observational skills. The birds usually sang and called during the trials, further facilitating our ability to locate them.

Having done occasional playbacks to banded wrens myself since 1971, I am sufficiently experienced with them and their habitat to know how impossible it would be for an observer to collect accurate data on the movements of birds as you describe in the paper. Now, in this response by Vehrencam et al., I am told there were three observers; that information is not available anywhere in Illes et al. And what is certainly not available is a description of how complicated movement data from three different observers are merged into one data file for analysis.

Third, the pseudoreplication criticism is a red herring. Each subject's stimulus exemplars were uniquely prepared from different base songs, and we used a wide variety of song types and source males for the base songs, thus eliminating the possibility of pseudoreplication.

Regarding pseudoreplication, I am simply going to say that I would not be so quick as to declare this matter a red herring. Using "a wide variety" of stimuli or sources isn't the kind of quantitative approach I'd use for determining whether I had psuedoreplicated or not. Here are some of the numbers I had gathered from your study, when I wrote the following in my Forum piece, and I still think they apply:

> *Playback stimuli from 25 males (consisting of 12 different trill types used to construct 19 different song types) are played back to 31 males; given that it is trill types that are manipulated, the simplest way to avoid pseudoreplication would be to use 12 trill types as the unit of analysis, not 19 song types, or 25 males, and certainly not 31 independent playbacks as was done in their analyses.*

Fourth, Kroodsma argues that the degree artificiality of manipulated playback stimuli could account for subject responses. Our modification of trill rate involved increasing or decreasing the silent gap between trill notes to a similar degree, so both alternative stimuli had equivalently altered note-to-interval ratios. The two final stimuli differed in trill rate by approximately 25–30%, so the modification represented a modest 10–15% change, i.e. they were not extremely artificial or abnormal songs. Even an individual banded wren may increase its trill rate by up to 7% during playback experiments compared to dawn chorus singing (Vehrencamp, Yantachka, Hall, & de Kort, 2013). The paired stimuli did have the same number of trill notes, and thus different durations. No experiment can perfectly control for all song variables. Banded wren songs vary greatly in duration both within and between song types and within and between individuals. Vehrencamp et al. (2014) found that longer songs were associated with more escalated contests. So any potentially confounding effect of song duration in the Illes et al. (2006) study would be conservative, since the theoretically higher performance (faster trill) stimulus had the shorter song duration.

It would be far more transparent for studies like this to disclose possible alternative explanations for their results. Inform the reader as to why the experiment did not "perfectly control for all song variables" and then let the reader, perhaps with some guidance from the authors, assess the implications of those variables that were not controlled. Not disclosing alternative explanations raises doubts as to what else the authors are not disclosing, and problems of overall credibility increase in magnitude.

Also, we humans cannot know how artificial or abnormal the manipulated songs sound to the birds.

Fifth, we did examine the tendency for trill performance components to vary in a consistent way among song types within males in the Vehrencamp et al. (2013) study, and we found largely consistent differences related to male age.

I was watching for some analysis that took this same data set and plotted on the graph the songs for different males and song types, to allow one to assess visually whether different males consistently fall on the graph in a way that would allow each to be assessed by his deviation scores. To say that there are “largely consistent differences related to male age” doesn’t get me to where I was hoping to get with this analysis. With males who are all the same age, for example, how reliably can one use “deviation” to assess the quality of a male? I would guess not at all. How consistent are the deviation values of a given male, and is there any hope that different males have consistently different deviations, such that deviations could be a reliable indicator of male quality? I would guess not at all.

Sixth, Kroodsma appears unable to consider that subjects that initially approached the fast stimulus would subsequently spend less time close to the speaker if the stimulus was a broad frequency bandwidth (low vocal deviation) song. Our conclusion for this result was that the subjects responded quickly to the more threatening stimulus, but were then more strongly repelled by the repeated playback of higher-performance trills. We (de Kort, Eldermire, Cramer, & Vehrencamp, 2009; Hall et al., 2006; Vehrencamp et al., 2007) and others (Collins, 2004; Searcy & Beecher, 2009) have written extensively about the difficulty of interpreting approach responses to alternative playback stimuli, and have recommended several solutions, such as presenting three alternative stimuli instead of two, monitoring other behavioural responses of receivers like singing and calling, and examining the sender's context and subsequent acts when delivering different song variants. The approach-negotiate-retreat sequence is typical of banded wren interactions (Vehrencamp et al., 2014). Rapid trill rate is an indicator of a highly motivated intruder (Vehrencamp et al., 2013), and a territory owner should respond to such a threat by approaching quickly (but not immediately attack). During close-range negotiation, repetitive delivery of the same song type indicates that a bird will no longer negotiate but will stand its ground (Molles, 2006; Vehrencamp et al., 2014). The Illes et al. study suggested that repetition of a broad bandwidth trill was especially threatening and caused the defending owner to back off after a shorter time. This repelling effect of broad-bandwidth songs was verified in a follow-up study by de Kort, Eldermire, Cramer et al. (2009), as discussed below. When birds back off, we know they are still interested in the stimulus because they keep singing, albeit from a distance. Thus Kroodsma's alternative explanation, that the birds were fleeing the slower, longer, low-performance songs, is inconsistent with the combined evidence from our other studies. Nuanced responses such as the one described in Illes et al. (2006) may be typical of two-speaker playback experiments to territorial male subjects (Reichert, 2011).

Here is what I wrote:

One strong summary statement requiring evaluation is the authors' conclusion that high-performance songs repel birds. The evidence for that statement seems to be as follows. (1) When the lower- and higher-performance songs are played simultaneously, birds are more likely to approach the higher-performance song first (13 of 17 birds, interpreted as an aggressive response towards the fast trill, high-performance song; P ¼ 0.049). (2) For subjects that approached within 10 m of a speaker, most (18 of 25) first approached the high-performance song, again considered an aggressive response towards the high-performance song (P ¼ 0.043 or 0.027). (3) For males approaching within 10 m of either stimulus, the time spent within 10 m did not differ for the lower- and higher-performance stimuli (P ¼ 0.182). (4) Time spent in a larger area near the two stimuli also did not differ (P ¼ 0.583). (5) In another analysis, however, the ‘16 males that entered the 10 m fast circle [where the fast-trill song was broadcast] at some point during the trial spent less time there the higher the performance score of their stimulus trill’ (P ¼ 0.020; page 1910). It is the initial strong, aggressive response towards the high-performance songs (items 1 and 2) and the subsequent reduced time spent within 10 m of higher performance songs (item 5), in spite of no differences between low- and high-performance songs when assessed simultaneously (items 3 and 4), that leads the authors to the following conclusion: ‘the subsequent decrease in aggressive response by the receiver suggests that the highest performance signals posed a threat so extreme that they effectively repelled rivals, even territory owners’ (page 1911). The logic used here is challenging to accept. Why would a bird first attack the more intimidating song, then subsequently be scared by it? How do the authors choose when and why and over what time frame a song should have what effect?

Readers will inevitably have two possible reactions to the decision tree that led to the conclusion that “*the highest performance signals posed a threat so extreme that they effectively repelled rivals, even territory owners.*”

Option 1: The authors know their birds so well that they know when the wrens will approach and how fast, when they’ll negotiate, retreat, stand their ground, be threatened, be repelled, back off, or respond in some other way that has been evaluated in their extensive studies. In spite of “the difficulty of interpreting approach responses to alternative playback stimuli,” the authors know best how to pick out which responses best reveal what the birds are actually doing. This approach would be more credible, of course, if the authors before their study had decided which response variables would be used, but it’s ok if they worked it out afterward.

Option 2: This is a typical example of how “researcher degrees of freedom” (Gelman) are used to generate a good story. In a study in which the birds’ responses could be documented in so many different ways, how on earth do you choose which particular set of numbers is going to be used to make the case for high performance songs repelling birds? *How do the authors choose when and why and over what time frame a song should have what effect?* I can’t help but believe that if you wanted to make the exact opposite conclusion, you could probably find the numbers for that conclusion, too.

Frankly, I am skeptical of what seems to me to be a highly convoluted, ad hoc explanation for how the birds responded to the playbacks. Given that the authors themselves find wide published support for the deviation/performance hypothesis of Podos, and I find none, their arguments to me are especially unconvincing. This is exactly the quality of support for the deviation/performance hypothesis that is so widespread, and I simply don’t buy it.

Finally, Kroodsma argues that we should have corrected all of the statistical tests in the entire results section with a Bonferroni multiple comparisons procedure. We think that a multiple comparison correction was not needed here. It is commonly acknowledged that the Bonferroni correction is far too harsh (i.e. Moran, 2003; Narum, 2006); the false discovery rate correction is superior in reducing type II errors, and we have done this correction in our papers where multiple variables were tested and presented in tables. In Illes et al. (2006), analyses were generated from three independent data sets that addressed completely different questions, thus they should not be combined as Kroodsma proposes. Some of the tests related to the playback experiment were presented to examine and dispel potential confounding effects. The remaining few tests addressed specific hypotheses and were not part of a multivariate fishing expedition to find the most significant variables. We presented power analyses and effect sizes for our tests, and these revealed stronger effects than the P values indicated.

I count eight statistical tests that were presented in the results for analyzing the responses of the birds to playbacks. I guess I have two reactions: 1) Those eight are independent tests and don’t need any correction for multiple comparisons? 2) If I had this data set, I would inevitably explore it in an untold number of ways to see what the birds were doing, and I’d have far done far more than the eight statistical tests that are presented.

As sceptical Sorry, but if one buys into the deviation/performance hypothesis of Podos and find it well-supported in the literature, I do not think you are sufficiently skeptical. scientists in search of the truth, we set out to further examine the interesting results in

Please cite this article in press as: Vehrencamp, S. L., et al., Response to Kroodsma's critique of banded wren song performance research, Animal Behaviour (2016), http://dx.doi.org/10.1016/j.anbehav.2016.12.004

Illes et al. (2006) with another playback experiment that manipulated only the frequency band- width of trills And just how normal were those manipulated songs to the birds? (de Kort, Eldermire, Cramer et al., 2009). This is a paper that I explicitly chose not to analyze in my forum, not because it was an exemplary piece of science, but because I thought I had done enough already. Contrary to Kroodsma's claim, this study was conducted with observers blind to the bandwidth treatments, Yes, good. Good for Cramer. and the differences could not be detected by the observers. We separately presented three alterna- tive bandwidth stimuli to subjects, and expressly quantified mul- tiple measures of male response to address the significance of nuanced retreat responses. The results strongly confirmed the

Please cite this article in press as: Vehrencamp, S. L., et al., Response to Kroodsma's critique of banded wren song performance research, Animal Behaviour (2016), http://dx.doi.org/10.1016/j.anbehav.2016.12.004

earlier study Were any alternative explanations explored for the results? : subjects avoided closely approaching high-performance stimuli the most abnormal of the stimuli? but continued to sing and call from a dis- tance; approached and negotiated with matching songs to the median-performance stimuli most normal songs?; and approached quickly but showed a lower vocal response to the low-performance stimuli abnormal again? (de Kort, Eldermire, Cramer et al., 2009).

I suppose I should go back and explore this deKort et al. paper, but I'm not going to do that. I have a hunch we'd have the same kinds of disagreements we have on Illes et al.

With a concern for how abnormal the manipulated songs are in these experiments, I repeat here a comment that I made in my Forum:

> *Rather than manipulating songs and adding multiple confounding variables, it would seem that a first, worthwhile experiment would simply compare responses to naturally occurring songs that are of high and low performance. Regardless of song type, on average, if performance matters the birds should respond differently to intact, unmanipulated songs at two extremes of performance.*

I'm not saying this is the perfect experiment, but it seems to me to be a logical place to start. No consistent response difference between high and low performance songs of the same male would suggest that the whole enterprise is not worth pursuing.

Another pair of experiments explored male responses to songs of different trill note consistency (de Kort, Eldermire, Valderrama, Botero, & Vehrencamp, 2009). One experiment used natural songs of the same type and from the same male in his first year versus in his second or third year, when males sing more consistently. The second experiment compared songs of first-year birds to the same song manipulated to have greater trill note consistency. Both experiments found stronger responses to the more consistent song stimuli. We emphasize that our use of natural songs here, as recommended by Kroodsma, produced similar results to the experiment with manipulated songs. As mentioned earlier, we showed in Vehrencamp et al. (2013) that trill note consistency of all measured song types increased in males from their first to their second and third year, and then plateaued or decreased slightly for older birds. Male age is not only associated with territorial defence experience, which could be assessed during territorial encounters, but multiple lines of evidence also suggested that females avoided mating with or divorced first-year males and preferred older and more consistently singing males as extrapair partners (Cramer, Hall, de Kort, Lovette, & Vehrencamp, 2011).

How consistently a male sings his song is altogether a different matter than the deviation/performance issue. Consistency has its roots in song development, with songs becoming more consistent as birds age, most noticeable to our ears during the first year of life, but as shown with these wrens, the consistency might increase as the birds age beyond their first year. This information is nice, but it is not directly relevant to the deviation/performance hypothesis of Podos.

Kroodsma's criticism of the language in the first sentence of the abstract of Vehrencamp et al. (2013) is totally unwarranted. We merely stated the general theoretical proposition being tested in our study, a standard protocol for scientific articles. The statement was fully justified given the large literature on performance constraints affecting sound production and the association of acoustic signal features with aspects of sender characteristics. Kroodsma does not appear to question the results, which showed that trill note consistency and frequency bandwidth increase with male age. We also discovered that trill note rate for a given song type increases during playback experiments in relation to the male's level of aggressive response, a result that has now been found in other species (Funghi, Cardoso, & Mota, 2015; Linhart, Jaska, Petruskova, Petrusek, & Fuchs, 2013). Thus this aspect of performance seems to provide cues to receivers about a rival's immediate aggressive motivation. We did not find any associations with male survival or our measure of body condition. Our results and interpretations were not biased by any desire to support or disprove the hypothesis, and in several instances we offered alternative hypotheses where appropriate.

Here are my "totally unwarranted" comments:

> *the opening sentence squarely places the context and rationale for this study in the realm of performance studies and sexual selection and honest signalling, with 'difficult-to-execute' sounds revealing male quality. Everything is interpreted in this context, yet there is no obvious scientific justification for doing so and good justification for not doing so (see especially the above review of Illes et al. (2006) on the same species). According to the scatterplot of trill rate and bandwidth for banded wrens (Fig. 11), relatively few songs are difficult to execute as defined in this performance context, because most songs fall far from the upper bound on the graph. Every male 'willingly' learns many 'low performance', easy-to-execute songs in order to have particular song types in his repertoire, as if performance did not matter, as if there were no selection for difficult-to-execute songs as claimed in this paper.*

I object to this opening because it manipulates readers to a point of view that I do not think is warranted, i.e., that birds assess one another based on the songs near the upper bound of the scatterplot because those songs plotting there are difficult to execute. I do not believe there is any evidence showing that birds care where a natural, manipulated song plots on the scatterplot. Songs might contain clues that tell the age of the singer, but nothing in the deviation tells of the relative quality of same-age males.

I have no reason to doubt that, as a male ages, songs might become more consistent, and I am not contesting those results. What I will claim is that there are no data showing that males assess one another based on "vocal deviation." i.e., a composite score of trill rate and frequency bandwidth.

If one wants to claim that songs approaching the upper bound are "difficult to execute," then why can't someone else claim that the songs approaching bounds on the graph to the left and bottom are also "difficult to execute." Why are those songs just abnormal and dismissed from further study?

I repeat what I have written numerous times before: I know of no credible evidence showing that the songs at the upper bound are interpreted by birds as "difficult to execute" and therefore "better," thus reflecting a better male (Ballentine et al. 2004),

Commenting further on this paper, Kroodsma (2017, page e14) writes (his italics): 'According to the scatterplot of trill rate and bandwidth for banded wrens (Fig. 11), relatively few songs are difficult to execute as defined in this performance context, because most songs fall far from the upper bound on the graph. Every male 'willingly' learns many 'low-performance', easy-to-execute songs in order to have particular song types in his repertoire, as if performance did not matter, as if there were no selection for difficult-to-execute songs as claimed in this paper'. Repertoire species such as the banded wren use contrasting song types to emphasize switching rates, short-term diversity and matching during territorial interactions (Molles, 2006; Vehrencamp et al., 2007, 2014). But Kroodsma has conveniently ignored another component of banded wren trills: their varied and complex note shapes as mentioned earlier. Trill note consistency is a third axis of performance in this species, and we showed in this paper (Vehrencamp et al., 2013, see supplementary online material) that consistency and vocal deviation trade off (are negatively correlated) within males and song types. Thus, song types far from the trill rate versus bandwidth upper limit are not necessarily easy to execute, as they may have a complex shape that is difficult to repeat consistently.

I'm tiring. Perhaps I simply ask if there are any data that could possibly be collected to falsify the deviation/performance hypothesis of Podos? I don't think so.

Kroodsma concludes that we still await good answers to the question of what information listeners extract about singers from their songs beyond species identification. In fact, there is a growing body of data showing that aspects of vocal performance, including trill rate, vocal deviation, frequency excursion, trill note and song consistency and call rate/call duration trade-offs, do provide useful information to receivers in some species and are associated with reproductive benefits in many birds, mammals, anurans and crickets (e.g. Botero et al., 2009; Byers, Akresh, & King, 2015; Funghi et al., 2015; Linhart et al., 2013; Pasch, George, Campbell, & Phelps, 2011; Petruskova et al., 2014; Podos et al., 2016; Reichert & Gerhardt, 2012; Sprau, Roth, Amrhein, & Naguib, 2013; Wagner, Beckers, Tolle, & Basolo, 2012; Welch, Smith, & Gerhardt, 2014). Our studies have contributed to this body of knowledge, specifically by demonstrating the existence of cues to age and aggressive motivation, along with the strategic use of song type use patterns to indicate approach and retreat during territorial negotiations.

It comes down to what you believe. How skeptical are you of published results? What exactly is the nature of that "growing body of data"? Throughout many of these studies (but not all; I highly respect some of these authors and their work), I see what Andrew Gelman writes about on his blog, and it's not good.

Acknowledgments

We thank Jeff Podos, Jack Bradbury, Dave Wilson and two anonymous referees for helpful feedback on earlier drafts of the manuscript. Our research was funded by the National Institutes of Health (R01-MH60461).

References

Ballentine, B., Hyman, J., & Nowicki, S. (2004). Vocal performance influences female response to male bird song: An experimental test. Behavioral Ecology, 15, 163e168.

Botero, C. A., Rossman, R. J., Caro, L. M., Stenzler, L. M., Lovette, I. J., de Kort, S. R., et al. (2009). Syllable type consistency is related to age, social status and reproductive success in the tropical mockingbird. Animal Behaviour, 77, 701e706.

Bradbury, J. W., & Vehrencamp, S. L. (2011). Principles of animal communication (2nd ed.). Sunderland, MA: Sinauer.

Burt, J. M., & Vehrencamp, S. L. (2005). Dawn chorus as an interactive communication network. In P. K. McGregor (Ed.), Animal communication networks (pp. 320e343). Cambridge, U.K.: Cambridge University Press.

Byers, B. E., Akresh, M. E., & King, D. I. (2015). A proxy of social mate choice in prairie warblers is correlated with consistent, rapid, low-pitched singing. Behavioral Ecology and Sociobiology, 69, 1275e1286.

Byers, B. E., & Kroodsma, D. E. (2009). Female mate choice and songbird song repertoires. Animal Behaviour, 77, 13e22.

ten Cate, C. (2004). Birdsong and evolution. In P. Marler, & H. Slabbekoorn (Eds.), Nature's music: The science of birdsong (pp. 296e317). San Diego, CA: Elsevier Academic Press.

Collins, S. A. (2004). Vocal fighting and flirting: The functions of birdsong. In P. Marler, & H. Slabbekoorn (Eds.), Nature's music: The science of birdsong (pp. 39e79). San Diego, CA: Elsevier Academic Press.

Collins, S. A., de Kort, S. R., Perez-Tris, J., & Telleria, J. L. (2011). Divergent sexual selection on birdsong: A reply to Byers. Animal Behaviour, 82(5), e4ee7.

Cramer, E. R. A., Hall, M. L., de Kort, S. R., Lovette, I. J., & Vehrencamp, S. L. (2011). Infrequent extra-pair paternity in the banded wren, a synchronously breeding tropical passerine. Condor, 113, 637e645.

Cramer, E. R. A., & Price, J. J. (2007). Red-winged blackbirds Ageliaus phoeniceus respond differently to song types with different performance levels. Journal of Avian Biology, 38, 122e127.

Funghi, C., Cardoso, G. C., & Mota, P. G. (2015). Increased syllable rate during aggressive singing in a bird with complex and fast song. Journal of Avian Biology, 46, 283e288.

Hall, M. L., Illes, A., & Vehrencamp, S. L. (2006). Overlapping signals in banded wrens: Long-term effects of prior experience on males and females. Behavioral Ecology, 17, 260e269.

Hebets, E. A., & Papaj, D. R. (2005). Complex signal function: Developing a framework of testable hypotheses. Behavioral Ecology and Sociobiology, 57, 197e214.

Illes, A. E., Hall, M. L., & Vehrencamp, S. L. (2006). Vocal performance influences male receiver response in the banded wren. Proceedings of the Royal Society B: Biological Sciences, 273, 1907e1912.

de Kort, S. R., Eldermire, E. R. B., Cramer, E. R. A., & Vehrencamp, S. L. (2009). The deterrent effect of bird song in territory defence. Behavioral Ecology, 20, 200e206.

de Kort, S. R., Eldermire, E. R. B., Valderrama, S., Botero, C. A., & Vehrencamp, S. L. (2009). Trill consistency is an age-related assessment signal in banded wrens. Proceedings of the Royal Society B: Biological Sciences, 276, 2315e2321.

Kroodsma, D. E. (2017). Birdsong performance studies: A contrary view. Animal Behaviour, 125, e1ee16.

Linhart, P., Jaska, P., Petruskova, T., Petrusek, A., & Fuchs, R. (2013). Being angry, singing fast? Signalling of aggressive motivation by syllable rate in a songbird with slow song. Behavioural Processes, 100, 139e145.

Logue, D. M., & Forstmeier, W. (2008). Constrained performance in a communication network: Implications for the function of song-type matching and for the evolution of multiple ornaments. American Naturalist, 172, 34e41.

Molles, L. E. (2006). Singing complexity of the banded wren (Thryothorus pleurostictus): Do switching rate and song-type diversity send different messages? Auk, 123, 991e1003.

Molles, L. E., & Vehrencamp, S. L. (1999). Repertoire size, repertoire overlap, and singing modes in the banded wren (Thryothorus pleurostictus). Auk, 116, 677e689.

Molles, L. E., & Vehrencamp, S. L. (2001). Neighbour recognition by resident males in the banded wren, Thryothorus pleurostictus, a tropical songbird with high song type sharing. Animal Behaviour, 61, 119e127.

Moran, M. D. (2003). Arguments for rejecting the sequential Bonferroni in ecological studies. Oikos, 100, 403e405.

Narum, S. R. (2006). Beyond Bonferroni: Less conservative analysis for conservation genetics. Conservation Genetics, 7, 783e787.

Pasch, B., George, A. S., Campbell, P., & Phelps, S. M. (2011). Androgen-dependent male vocal performance influences female preference in Neotropical singing mice. Animal Behaviour, 82, 177e183.

Petruskova, T., Kinstova, A., Pisvejcova, I., Mula Laguna, J., Cortezon, A., Brinke, T., et al. (2014). Variation in trill characteristics in tree pipit songs: Different trills for different use? Ethology, 120, 586e597.

Podos, J. (1997). A performance constraint on the evolution of trilled vocalizations in a songbird family (Passeriformes: Emberizidae). Evolution, 51, 537e551.

Podos, J. (2017). Birdsong performance studies: Reports of their death have been greatly exaggerated. Animal Behaviour, 125, e17ee24.

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.

Reichert, M. S. (2011). Effects of multiple-speaker playbacks on aggressive calling behavior in the treefrog Dendropsophus ebraccatus. Behavioral Ecology and Sociobiology, 65, 1739e1751.

Reichert, M. S., & Gerhardt, H. C. (2012). Trade-offs and upper limits to signal performance during close-range vocal competition in gray tree frogs Hyla versicolor. American Naturalist, 180, 425e437.

Searcy, W. A., & Beecher, M. D. (2009). Song as an aggressive signal in songbirds. Animal Behaviour, 78, 1281e1292.

Sprau, P., Roth, T., Amrhein, V., & Naguib, M. (2013). The predictive value of trill performance in a large repertoire songbird, the nightingale Luscinia megarhynchos. Journal of Avian Biology, 44, 567e574.

Trillo, P. A., & Vehrencamp, S. L. (2005). Song types and their structural features are associated with specific contexts in the banded wren. Animal Behaviour, 70, 921e935.

Vehrencamp, S. L. (2001). Is song-type matching a conventional signal of aggressive intentions? Proceedings of the Royal Society B: Biological Sciences, 268, 1637e1642.

Vehrencamp, S. L., Ellis, J. M., Cropp, B. F., & Koltz, J. M. (2014). Negotiation of territorial boundaries in a songbird. Behavioral Ecology, 25, 1436e1450.

Vehrencamp, S. L., Hall, M. L., Bohman, E. R., Depeine, C. D., & Dalziell, A. H. (2007). Song matching, overlapping, and switching in the banded wren: The sender's perspective. Behavioral Ecology, 18, 849e859.

Vehrencamp, S. L., Yantachka, J., Hall, M. L., & de Kort, S. R. (2013). Trill performance components vary with age, season, and motivation in the banded wren. Behavioral Ecology and Sociobiology, 67, 409e419.

Wagner, W. E., Jr., Beckers, O. M., Tolle, A. E., & Basolo, A. L. (2012). Tradeoffs limit the evolution of male traits that are attractive to females. Proceedings of the Royal Society B: Biological Sciences, 279, 2899e2906.

Welch, A. M., Smith, M. J., & Gerhardt, H. C. (2014). A multivariate analysis of genetic variation in the advertisement call of the gray treefrog, Hyla versicolor. Evolution, 68, 1629e1639.

Wells, K. D., & Taigen, T. L. (1986). The effect of social interactions on calling energetics in the gray treefrog (Hyla versicolor). Behavioral Ecology and Sociobiology, 19, 9e18.