Animal Behaviour xxx (2016) e1ee8

Contents lists available at ScienceDirect

Animal Behaviour

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

Forum

Birdsong performance studies: reports of their death have been greatly exaggerated

Jeffrey Podos*

Department of Biology, Graduate Program in Organismic & Evolutionary Biology, University of Massachusetts, Amherst, MA, U.S.A.

a r t i c l e i n f o

Article history:
Received 20 July 2016
Initial acceptance 19 November 2016
Final acceptance 1 December 2016
Available online xxx
MS. number: AF-16-00647

Keywords:
frequency bandwidth
sexual selection
song
trill rate
vocal performance

Kroodsma (2017, Animal Behaviour, 125, e1ee16) has critiqued 'the performance hypothesis', which posits that two song attributes, trill rate and frequency bandwidth, provide reliable indicators of singer quality and are used as such in mate or rival assessment. Kroodsma develops three main arguments: (1) young male sparrows copy songs with high accuracy from neighbours, and thus cannot calibrate song models to their own performance capacities; (2) in species with song repertoires, vocal performance varies widely within individuals and among song types, thus rendering song performances inadequate as quality indicators; and (3) experimental studies of song function have relied on playback of structurally abnormal stimuli, with interpretation of birds' responses to these stimuli thus compromised. I address these critiques in turn, offering the following counterpoints: (1) the reviewed literature actually reveals substantial plasticity in song learning, leaving room for birds to tailor songs to their own performance capacities; (2) reasonable scenarios, largely untested, remain to explain how songs of repertoire species could convey information about singer quality; and (3) the playback studies critiqued actually enable direct, reasonable inferences about the function of vocal performance variations, because they directly contrast birds' responses to low- versus high-performance stimuli. My analyses support the plausibility of performance hypotheses and highlight avenues for future research. My analyses also reveal numerous shortcomings with Kroodsma's arguments, including an inaccurate portrayal throughout of publications under review, logic that is thus rendered questionable and reliance on original data sets that are incomplete and thus inconclusive.
I. e., There's "room" with "reasonable scenarios" so it's all "plausible" . . .

(My text in yellow hi-lite throughout—

for more details, see http://donaldkroodsma.com/?page_id=1596)

My title:

This Birdsong Performance Literature* defended by Podos Thrives on Half-Truths, Deception, and Abundant 'Researcher Degrees of Freedom'

**more specifically, the literature that relies on the scatterplot of frequency bandwidth and trill rate, and the distance that a given song plots from an upper bound, i.e., "deviation" (and related literature)*

Please cite this article in press as: Podos, J., Birdsong performance studies: reports of their death have been greatly exaggerated, Animal Behaviour (2016), http://dx.doi.org/10.1016/j.anbehav.2016.12.010

e**1) My abstract: The particular birdsong performance literature that I critique[1], and that Podos defends here, is fiction.[2]**

The considerable literature that has developed on this topic is a textbook example of how highly flawed, indefensible methods have been used by prolific, influential authors to promote intuitively appealing but false stories, all of which has been accepted uncritically by others and become entrenched in the literature. This literature thrives on half-truths (or half-lies), *never telling the whole truth*. The most obvious deception is the Rule of Consistency: As long as some data can be found to be *consistent with* a favored idea, no alternative explanations need be considered or mentioned, especially mundane explanations that do not make for a good story. [3] Other deceptions fall under the general category of "researcher degrees of freedom," offering the authors so much flexibility in collecting, analyzing, and presenting their works that they are able to make "promotable claims about 'the pervasiveness and persistence of . . .' whatever . . . they want to publish that day" (I offer abundant examples throughout this document as I respond to Podos' defense of this literature)."[4] The same problems occur throughout other literature on birdsong and sexual selection, which for shorthand here I simply refer to as the "large song repertoires impress" and the "nutritional stress" ideas. The long-term damage by these studies to a scientific study of birdsong and sexual selection has been immense.

Podos' solution to all of this bad news is to shoot the messenger: "take anything Kroodsma has published, critiques and science alike, with a heavy dose of skepticism . . ."[5] (Questioning can begin at http://donaldkroodsma.com/.) My solution is to impose (very) heavy costs on

---

[1] Using "deviation" from an upper bound on a scatterplot of trill rate and frequency bandwidth as a measure of how difficult it is to produce a bird song, and as a measure of the quality of the male who produced the song . . . and related literature.

[2] Merriam Webster online definition for kids: "**1)** something told or written that is not fact; **2)** a made-up story," i.e., using all available "researcher degrees of freedom" (Gelman) to generate a good story, i.e., there is no truth to these stories.

[3] For example, a song is routinely manipulated so that it is, by some chosen measures, considered to be "low performance" or "high performance," though it is also highly abnormal. If a territorial male does not vigorously attack this song, as he would a normal song, the only interpretation offered is that he is so intimidated by this song that he flees, because such a high-performance song must surely reflect the high quality of the frightful male who delivered it. No mention is made of the likely (unexciting, unpublishable) explanation that the song is so abnormal that it has no meaning for the territorial male.

[4] Quotes from Andrew Gelman's blogs, such as this one: http://andrewgelman.com/2016/09/21/what-has-happened-down-here-is-the-winds-have-changed/. Other problems popularized by Gelman, with which this birdsong performance literature is replete, include p-hacking, research incumbency rules, the garden of forking paths, vampirical theory, multiple comparisons problems, " . . . the find-statistical-signficance-any-way-you-can-and-declare-victory paradigm," and more. I would add the Rule of Consistency: "Find-anything-consistent-with-whatever-you-want-to-publish-that-day-and-declare-victory-and-think-no-more."

[5] Podos quote, from Andew Gelman's blog (http://andrewgelman.com/2017/08/13/bird-fight/): "Stepping back for a moment, and as a bit of a public service announcement, Kroodsma's stubborn adherence to contrarian viewpoints here comes as no surprise to veterans of our field. As Vehrencamp et al. (2017) observed, "Kroodsma has previously applied the same modus operandi in attempts to discredit other major topics in the field of avian vocal communication. Of course there's nothing wrong with reasonable, fair criticism, as noted eloquently by Vehrencamp et al. in their opening paragraph. Yet Kroodsma's critiques have moved well beyond reasonable and fair. So I offer a message to future students stumbling on this thread; take anything Kroodsma has published, critiques and science alike, with a heavy dose of skepticism . . ." . . . (My response to Vehrencamp et al.'s "eloquent" opening paragraph is also on my website, http://donaldkroodsma.com/?page_id=1596, and addressed briefly below as well.)

Please cite this article in press as: Podos, J., Birdsong performance studies: reports of their death have been greatly exaggerated, Animal Behaviour (2016), http://dx.doi.org/10.1016/j.anbehav.2016.12.010

authors who create, publish, promote, and defend this kind of literature, so as to return the study of birdsong and sexual selection to the realm of science.

There comes a time for frank talk, and this is it, lest anyone think that the disagreements that Podos et al. and I have are minor issues about how one does science. Herein is a candid, no-holds-barred, no-punches-pulled response to Podos' defense of this literature.

Contents

Please cite this article in press as: Podos, J., Birdsong performance studies: reports of their death have been greatly exaggerated, Animal Behaviour (2016), http://dx.doi.org/10.1016/j.anbehav.2016.12.010

Please cite this article in press as: Podos, J., Birdsong performance studies: reports of their death have been greatly exaggerated, Animal Behaviour (2016), http://dx.doi.org/10.1016/j.anbehav.2016.12.010

2) Prologue. How the hell could all this have happened?

The short term, since 2014:

Absent from scientific meetings for a decade, blissfully enjoying a retirement from academic life, I attended an ornithological meeting in 2014 and heard a paper that totally floored me, both because it had to be totally false and because it won a best student paper award (Goodwin and Podos 2014). I snooped some more, and soon discovered a burgeoning literature of which this paper was just the culmination. The authors refused to communicate with me, and threatened me with criminal harassment charges, by way of the UMass police, for asking about their research, thus preventing me from talking to someone in my own Biology Department at the University of Massachusetts. Attempts at a dialogue in *Biology Letters* were halted by a secret letter from the University of Massachusetts, sent by Podos and supposedly written by the dean of the graduate school, who admitted later to knowing nothing about the letter. Initial attempts to publish a Forum article in *Animal Behavior* were halted abruptly when the target authors convinced the editor (a former editor) to reject my article (in an angry, public letter) before it was even submitted. Attempts to address matters of scientific and ethical misconduct with UMass university administrators were dismissed. . . . Eventually, another editor at Animal Behavior saw every reason to proceed with a normal submission process. Referees were supportive. Eventually my Forum was published, and Podos, Vehrencamp et al., and Cardoso responded. I replied to them. Over more than three years, I am aghast at where this investigation has taken me (all documented at http://donaldkroodsma.com/?page_id=1596).

I hope to be finished with this odyssey now. The big questions are . . . Who cares? Will the quality of published work on birdsong and sexual selection be improved? Will the costs of producing and defending this work finally outweigh the benefits?

The long term:

How has such a flawed literature flourished for so long? Good question. I don't have any easy answers to this question. 1) One reason is that critical reviewers like me are never asked to assess this work for publication or grant support. Ever since my signed, highly negative review of a Podos paper in 2004, in which I said he was more interested in a good story than the truth (marketing vs. science), I have never seen another. The same for others in this field. I have always been curious what Podos and his colleagues say to journal editors or granting agencies about my being inappropriate as a reviewer. 2+) There are plenty of other reasons as well, reasons that are sociological, etc. Overall, the fundamental reason this literature has flourished is that the benefits have always outweighed the costs. And that's a pretty sad commentary on lots of things.

Please cite this article in press as: Podos, J., Birdsong performance studies: reports of their death have been greatly exaggerated, Animal Behaviour (2016), http://dx.doi.org/10.1016/j.anbehav.2016.12.010

3) I wasn't going to do this, but Jeff Podos asked for it (literally)

On Andrew Gelman's blog (http://andrewgelman.com/2017/08/13/bird-fight/), Jeff Podos has expressed his considerable displeasure not only with my original Forum article in Animal Behavior (Kroodsma 2017) but also my anemic (but perhaps civil and therefore publishable) response to this document by Podos, which is his response to my Forum article. It's a confusing sequence of events, but there it is: I write a Forum article (A1), Podos (and others) reply to it (B1), then I reply to their replies (A2), then maybe they get to reply to my reply (B2). My reply to their reply is available but not published. All of these documents are accumulating on my website (http://donaldkroodsma.com/?page_id=1596), where the sequence of events is I hope clear.

On Andrew Gelman's blog (http://andrewgelman.com/2017/08/13/bird-fight/), Podos advises the reader to "ask yourself: has Kroodsma responded to any of the counterpoints raised both by me and Vehrencamp et al.? If so, has he been able to reconcile his positions with the actual published literature?"

I have two reactions to that quote:

First, I have no respect for the "actual published literature" on this deviation/performance literature, so there's nothing respectable to reconcile with (Andrew Gelman refers to what he's " . . . called the research incumbency rule: that, once an article is published in some approved venue, it should be taken as truth . . ."—I do not subscribe to this rule).

Second, no, I have not responded to the "counterpoints," because I felt they were so far off the mark that I thought I'd let them stand on their own, that to respond in detail would give them an air of legitimacy.

Podos has asked for a response, so I will do so here, to reveal candidly, in detail, what is so wrong about this literature I have critiqued. What I find so remarkable is that Podos and I find the other's stance so untenable. I might suggest that perhaps he and I could have talked through some of this, face to face, had he not hidden behind the threat of criminal harassment charges if I asked questions about his work. My entire Forum article probably would not have happened if Podos had been more forthcoming about how he does his research. But probably I'm dreaming. He's refused to speak to me since 2004 when I told him he was marketing stories rather than doing science, and it's only gotten worse since then.

To proceed . . . Yes, contrary to Podos' suggestion on the Gelman blog, I had studied his response to my Forum, and mentally noted not a few objections to what he wrote (same for Vehrencamp). So here, in Podos' document converted from pdf to Word, I will respond in the kind of detail that he has asked for. I regret that I am doing this, but it seems that if Podos doesn't understand the points in my Forum, or if he is intentionally trying to create doubt where none should exist, others will also need

[e]explicit commentary or be will be deceived by Podos' response. My comments will be hi-lited in yellow, the original text on which I am commenting in red.

I am going to try to limit my reply mostly to sections on chipping sparrows and let it serve as a proxy for the rest. That does not mean that I could not respond to the rest if pressed, i.e., explanation # 1 above does not apply anywhere. If I felt this were an honest exchange about science and how we might advance our understanding of birdsong, I'd be all in and comment more thoroughly. But I don't believe that's what this is all about, as has been clear to me now for several years (see my website for details of secrecy that have shrouded the research by Podos and students).

4) Outrage

And, at times, I may be uncivil and intemperate, using words that I feel are appropriate, as I believe that not a small measure of outrage is deserved for what has transpired in this literature and for Podos' defense of it.

Please cite this article in press as: Podos, J., Birdsong performance studies: reports of their death have been greatly exaggerated, Animal Behaviour (2016), http://dx.doi.org/10.1016/j.anbehav.2016.12.010

In his critique of literature on vocal performance in birds, Kroodsma (2017) adopts positions that run the gamut from healthy sceptic to full-throated contrarian. My goals here are to distill and evaluate the main critiques presented and to highlight areas for future work.

As a preliminary comment: the focal topic here is what Kroodsma refers to as 'the performance hypothesis'. Rather than framing a single hypothesis, I find it clearer to parse the relevant content into two smaller-scale hypotheses, each of which has its own history, conceptual bases and methods for study. The first addresses vocal learning and production, and asks what role performance constraints might play in shaping song structure. Performance constraints are indeed recognized as influencing wide-ranging song features (e.g. Cardoso, Atwell, Ketterson, & Price, 2007; Lambrechts, 1996; Pasch, George, Campbell, & Phelps, 2011; Reichert & Gerhardt, 2012; Sakata & Vehrencamp, 2012; Suthers, Vallet, & Kreutzer, 2012; Zollinger & Suthers, 2004; reviewed by: Podos, Lahti, & Moseley, 2009; Podos & Patek, 2015), and it follows that low-quality individuals might encounter particular difficulties in producing song features that are performance-limited (e.g. Cardoso, 2013; Johnstone, 1997; Searcy & Nowicki, 2005). For the two features in question, trill rate and frequency bandwidth, the first hypothesis (H1) can be stated as follows. It is more difficult, because of performance challenges, to develop and sing trills with faster trill rates and/or wider frequency bandwidths. Trill rate and frequency bandwidth values can thus provide reliable indicators of singer quality.

The second smaller-scale hypothesis addresses song perception and function, and asks whether animals listening to song are able to discriminate performance-related variations and, if so, whether they modulate their behaviour in accordance. A large body of work indicates that animals of many species do indeed attend and respond differentially to performance-related vocal variations (e.g. Byers, 2007; Forstmeier, Kempenaers, Meyer, & Leisler, 2002; Geberzahn & Aubin, 2014; Welch, Semlitsch, & Gerhardt, 1998;

* Correspondence: J. Podos, Department of Biology, University of Massachusetts, Amherst, MA 01003, U.S.A.
E-mail address: jpodos@bio.umass.edu.

[Column and page breaks are retained, more or less, as in the original, often giving breaks at odd places in this "edited" version.]

http://dx.doi.org/10.1016/j.anbehav.2016.12.010

reviewed by Podos et al., 2009). For the two features in question, the second hypothesis (H2) can be stated as such: Animals discriminate and respond differentially to songs with varying trill rates and frequency bandwidths, in directions consistent with sexual selection theory.

And consistent with a whole bunch of other explanations too. Already the carefully chosen words have shifted away from "deviation," which is the supposed tradeoff between trill rates and frequency bandwidths.

SONG LEARNING: ACCURACY VERSUS PLASTICITY

The first of Kroodsma's primary critiques might be summarized as follows. (1) The performance hypothesis posits that birds' songs reflect individual variation in their production capacities, with higher-quality singers able to learn and produce songs with faster trill rates and/or wider frequency bandwidths (my H1 above). (2) Yet, available evidence indicates that birds copy their songs from adult tutors with striking accuracy. (3) This premium on learning accuracy precludes opportunities for individual birds to fine-tune songs to their own performance capacities. (4) Therefore, the performance hypothesis is untenable. Kroodsma develops this argument first for chipping sparrows, Spizella passerina, and then for swamp sparrows, Melospiza georgiana.

Chipping Sparrows

Again, most of my response will be to chipping sparrows, though I will not be able to resist comment on particularly egregious passages elsewhere. [Returning here after a first pass through the document, I realize that wasn't very good at resisting.]

For chipping sparrows, Kroodsma's argument builds on Liu and Kroodsma (2006), a descriptive study of dispersal and song learning. The main findings of Liu and Kroodsma (2006) were that yearling chipping sparrows learn to sing by copying a single adult tutor from a territory adjacent to their own, and that this copying tends to be very precise. The map in Liu and Kroodsma (2006), reprinted by Kroodsma (2017, his Figure 1), indeed shows cases of young males, having recently settled on a territory, singing the same song type as an adult neighbour. Moreover, Kroodsma presents analyses of original recordings that show clusters of songs with shared structure. Both lines of evidence seem to support Kroodsma's thesis: how could song structure in chipping sparrows reflect anything but accurate learning, as opposed to birds' individual performance capacities?

5) The value of descriptive research

Two comments.. I cherish that word "descriptive," in a way that few people do these days. "To experiment first is human, to describe first divine," a coauthor (Bruce Byers) and I wrote a few decades ago. Few journal editors or referees want "just a descriptive paper." No, give us experiments that are on the cutting edge of this or that. As I write in my counterreply to Podos, "*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*."

Second, . . . I'm going to let this go for now . . . except to say that . . . It's a clever setup for what is to follow, and many readers will form their opinion right here, unless they risk reading more.

I offer two observations in counterpoint. First, Liu and Kroodsma's (2006) study not only illustrates cases of accurate copying, but also shows that young male chipping sparrows, as they learn to sing, typically have the chance to interact with multiple singing neighbours. The map in Liu and Kroodsma (2006; again see Kroodsma, 2017, his Figure 1) illustrates this point nicely: males pack their territories tightly, most individual males have multiple neighbours, and most males in an area sing song types that are highly distinct from one another. Furthermore, within this spatial context, young males appear flexible as to which of their neighbours' songs they will copy. To quote Liu and Kroodsma (2006, page 516; see also Liu & Nottebohm, 2007):

> [j]uvenile Chipping Sparrows produce 5e7 'precursor' song syllables that encompass the acoustic space of species-specific adult songs. These precursor syllables are different from adult song syllables and are not used during the breeding season, but they remain plastic for later modification. In spring, upon hearing a new song, a juvenile can rapidly (within a few days) modify one of his precursor syllables to perfectly match the new song...The precursor songs do not rely on imitation...but the young birds do require auditory feedback to fully express these precursor repertoires. This learning mechanism ensures species identity yet allows enough flexibility for song acquisition under a variety of circumstances...

It is but a small step to posit that a young male chipping sparrow will choose which neighbour to copy based on experience with his own performance capacities, as he progresses through song learning's sensorimotor phase (Podos et al., 2009).

6) The preposterous proposition that a male chipping sparrow chooses his trill rate to match his unique vocal proficiency

Many small steps are fatal, and this one is not so small. Yes, of course a chipping sparrow can hear many different songs, and of course he can choose which one to learn to sing, based on something that's happening in his own head, based on territorial interactions with the tutor, during his first fall or the following spring, or whatever/whenever.

Now let's look carefully at Podos' proposal that a male chooses which trill rate to sing so as to match his own performance ability. Try staring briefly at Figure 4 (copied below) from Goodwin and Podos (2014), with a few of my embellishments added to the figure. It's easy to gloss over the inconvenient details in the word game that Podos plays with readers, but not so easy when staring at some cold facts.

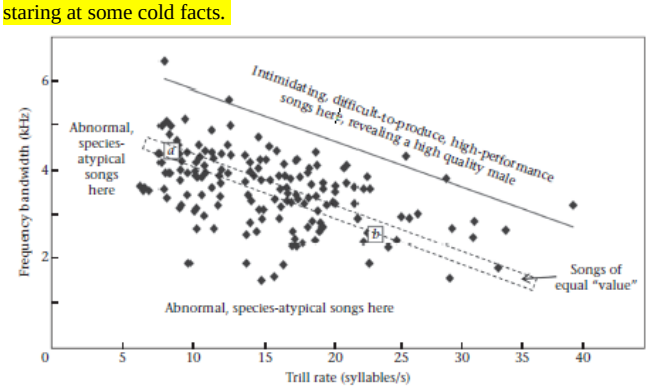

Figure 4. Reprinted with permission from Figure 1 in Goodwin and Podos (2014) . ' Chipping sparrow songs show evidence of a vocal constraint ... Biplot of trill rate and frequency bandwidth (N = 160 males) reveals a performance trade-off in vocal production ' . Letters ' a ' and ' b ' refer to a portion of the original figure not illustrated here. Songs that plot within the dashed rectangle have similar deviations from the upper bound and would therefore be considered equally good, or proficient, in other performance studies. Data are replotted on expanded axes to show the open space below and to the left of the data points.

Here's one of the first facts that I see: The natural range of trill rates among chipping sparrows is roughly 7 to 40. That's a huge range.

Let's be explicit: Podos is claiming that a male with a trill rate of 7 is far inferior to one with a trill rate of 40. He is claiming that a male listens to the range of songs available to him and then settles on learning a song with a trill rate that matches his own vocal proficiency.

How can this work on an evolutionary scale. How can such a range of variability be maintained in the population, if the birds are assessed based on those very songs? Males in the, what, lower nine tenths of the trill rate spectrum are complete losers? Males with a trill rate slower than 30 need not show up in the mating competition? All females would flock to males on the high end? One could predict the extra-pair copulation rates within a population, as you'd expect the males with faster trill rates to fertilize more females, if the females knew what they should be doing if they read the literature. I suppose a "scenario" could be found to accomplish all this, as scenarios and hypotheses and guesses form a lot of the basis for what the birds might be doing.

As an ornithologist who reads and publishes (or, rather, has published—past tense) primarily in the ornithological literature, I find the above all rather, well, preposterous. According to Podos, where a male settles on the trill rate spectrum (x-axis) depends on his vocal proficiency. Try getting that published in the ornithological literature. I don't believe Podos would consider himself an ornithologist, as when I last looked he was an author of only one paper in an ornithological journal (and that as a distant author in the journal *Emu*). That's not to say that nonornithological journals publish only rubbish, of course (but I can see where my statement could be taken)! What I am saying quite explicitly is that a referee who knew birds and specifically the biology of chipping sparrows would never have recommended publishing the paper by Goodwin and Podos (2014), for additional reasons outlined in my Forum and repeated below.

The default explanation for an ornithologist is that each of these trill rates, each of the frequency bandwidths, i.e., each of these songs falling within the normal range of chipping sparrow songs, is likely to be equally viable in the population (take the statement within reason, not to an extreme). If it were otherwise, the selective consequences would be severe.

7) Failure to disclose simple, base-line, default explanations for data is misleading:

And here is one of my main beefs for all of the literature that Podos defends: Rarely are the simple, base-line, default explanations for data offered as a viable explanation. Those explanations fail to get mentioned, conveniently making for a much better story. (More on this later.)

This scenario differs somewhat from that suggested by Kroodsma (2017, page e2), who concedes that 'One might argue, if pressed, that a young male could innately know his relative singing ability and then choose to settle next to an adult whose song he can master'. In my reading of Liu and Kroodsma (2006), the more likely scenario is that a young bird would not require an a priori sense of his own performance capacities at all, nor need to make an overt decision to settle next to a male tutor whose songs he could master. Rather, a young male might settle anywhere, begin to discover his vocal proficiency as he learns to sing, and then choose the tutor from among his neighbours whose song best matches his own vocal proficiency.

I just erased a longish comment about this text. Enough is enough.

My second observation concerns plasticity of individual song types themselves. Working in the laboratory with hand-reared male chipping sparrows, Liu and Kroodsma (1999) tracked song ontogeny in a cohort of males and reported a diverse array of plastic learning strategies. One juvenile male (JM1) originally matched his trill rate to that of its tutor, but then later increased his trill rate, during the plastic stage, by about 35%. Another male (JM8) started singing one song type but then switched suddenly, without any detected transition period, to sing another type. Yet another male (JM3) sang a song that initially mimicked a hatching-year tutor but then modified his song to more closely match a spring tutor. Other similar such examples are presented. It is this kind of plasticity that likely generates, over evolutionary time, the kind of structural diversity we see across song types in this species.

Thus, young male chipping sparrows seem to be exceptionally active vocal learners (Marler, 1997), able to alter both the choice of song type they copy and the acoustic structure of those types. This evidence diametrically opposes (Really? "diametrically opposes"? See Figure 2 below) Kroodsma's principal assertion that 'all features of a male's song, including his trill rate…are determined by [his] adult tutor' (Kroodsma, 2017, his Figure 2 caption, page e3). To the contrary, the flexible nature of chipping sparrows' learning programme would seem to allow ample opportunities for young birds to shape songs to their own individual performance capacities.

8) Deceptive wording: The focus must remain on "trill rate," not shift to general "performance capacities"

Of course a young male can learn any song he chooses to. It's a matter of which song he "wants" to learn. Like probably all songbirds, he as a young bird "overproduces," producing a rudimentary attempt at a variety of songs that he might have been exposed to. But to suggest that he's deciding where to sing on the trill rate spectrum of 7 to 40 is, well, a pretty puzzling, highly novel proposition. It's a pie-in-the-sky explanation, ***without mention*** of the basic biology of how a chipping sparrow learns his songs.

Look at those last two enlarged words in Podos' paragraph above. He has now shifted the wording to some general "performance capacities" from what this dialogue is all about, which is trill rate. Goodwin and Podos (2014) is about trill rate, not some general performance capacities (except when it's not). More on this below.

To illustrate this point about "trill rates" with an example, let me show what Podos is proposing happens in nature. Below is Figure 2 from my Forum, with a self-explanatory legend. Podos is claiming that the songs on the far right, with a trill rate of 25, are "better" than the songs on the far left, with a trill rate of 7. If he is capable, Podos is arguing, a young male should choose song type #14 instead of song type #1, because a male who sings song type #14 reveals a superior vocal proficiency, and then all kinds of good things will happen for him. The males who sing song type #1 are losers, those who sing song type 14 exceptional.

In my original Forum article, I point out that "*there is no evidence for song learning in any songbird species or especially in chipping sparrows (Liu & Kroodsma, 1999, 2006) that a male is in any way limited in what naturally occurring trill rate he can learn,"* and certainly nothing of the magnitude that Podos proposes.

The reader might also take a look at Podos' "diametrically opposes" comment when viewing this figure. It would appear that a young male learns the details of his songs (trill rate, frequency range, etc.) from an adult tutor. Yeah, I know, I cannot claim that these pairs of songs are actually from an adult tutor and his tutee (a distracting point that Podos raises somewhere), but they look exactly like actual examples as documented by following many birds in nature.

Please cite this article in press as: Podos, J., Birdsong performance studies: reports of their death have been greatly exaggerated, Animal Behaviour (2016), http://dx.doi.org/10.1016/j.anbehav.2016.12.010

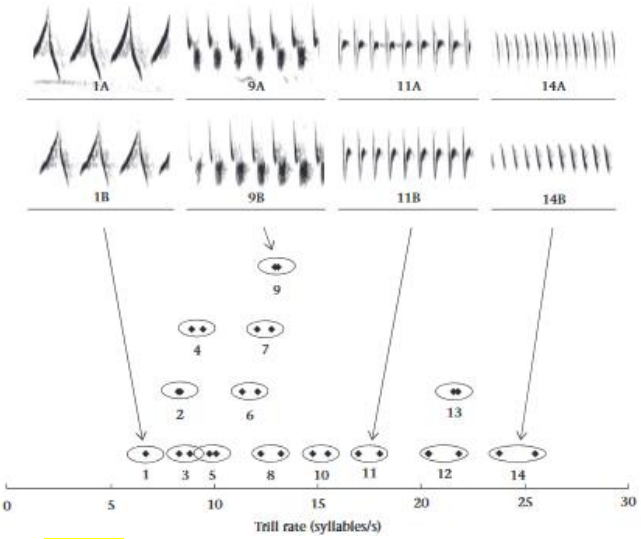

Figure 2.
A few dozen different song types can occur within a chipping sparrow population (only four illustrated here: 1, 9, 11, 14), but neighbouring males (A an d B) often have nearly identical songs, the result of a young male copying the song of a nearby adult singer (Liu & Kroodsma, 1999, 2006);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

9)Yes, "we still have much to learn". Who's going to do it, with objectivity and credibility?

To be clear, we still have much to learn about whether and how

The "we still have much to learn" argument is a distraction from the current work that is presented on chipping sparrows in Goodwin and Podos (2014). Of course "we still have much to learn." It was the same concluding argument used in Goodwin and Podos (2015) in their rebuttal to Ackay and Beecher (2015): "Nothing to see here move along" let's focus on all the good work we can do next. As I say in my article to be published in Animal Behavior, there is much work to be done, but I have no confidence that Podos is the one to oversee this work.

song development in this species might actually reflect individual variation in performance capacities. That would require a focused study of individual variation in male quality during sensorimotor ontogeny, and its effects on song motor development. Podos et al. (2009, pp. 181e182) outline an experimental approach to achieve that goal, which would involve manipulating individual condition using a developmental stress paradigm, and then tracking potential effects on the development of performance-related song features. In the meanwhile, it is surely premature to conclude that the available learning data on chipping sparrows (reviewed above) 'seriously undermine' the conclusions of Goodwin and Podos (2014), specifically with regard to the implication that performance-related song variations in chipping sparrows reflect individual variation in singer quality.

Absolutely, how chipping sparrows learn their songs from such a broad range of trill rates seriously undermines the entire Goodwin and Podos argument. In fact, I suggest that it is so serious that the authors conveniently forgot to mention the known facts about how birds in nature acquire their songs. There is no excuse for hiding this information from readers. To hide that information so as to make an argument about how trill rates affect coalitions is beyond . . .well, I lose words here. It's simply highly deceptive and has no place in scientific literature. It makes for a good story, but that's not science.

10) More deceptive wording: "performance" vs. trill rate (again)

"it is surely premature . . .specifically with regard to the implication that performance-related song variations in chipping sparrows reflect individual variation in singer quality" Very slippery. The paper by Goodwin and Podos (2014) is not about some general performance-related song variations, but instead specifically about trill rates. Of course there is more to learn about how chipping sparrows convey something about their ability by their song "performance." That is not the issue, but instead a clever distraction. The issue is whether TRILL RATE can be a performance measure that the birds use. I say it is preposterous to even suggest that, especially given the raw facts in the figures illustrated above and how the birds learn their songs.

11) "The original wording had been chosen with care"—implications of that statement

As Podos writes later in this document, "the original wording had been chosen with care." Yes, these words about "performance-related song variations" were no doubt carefully chosen, too, and they mislead the reader and change the subject entirely. This is deception, period.

12) Here's the clincher: All of the above is merely a distraction from this one key issue, here:

And one realizes that all this is even more outlandish when the bigger picture is considered. Look again at figure 4 above. The performance/deviation hypothesis is all about tradeoffs, this being between TRILL RATE and FREQUENCY BANDWIDTH. Neither by itself has been proposed to be an indicator of male quality; rather, it is the tradeoff between the two, so that a song that plots closer to the upper bound is "more difficult" to sing and therefore "better" because it indicates a higher quality male, because *the combination of fast trill rate or broad frequency* plots it closer to the upper bound. One could draw a longish rectangle encompassing all songs of equal value, as I did on the original graph of Goodwin and Podos (2014). No one would expect trill rate alone or frequency bandwidth alone to reflect male quality. No one, at least until now, as in Goodwin and Podos (2014), and strongly defended by Goodwin and Podos (2015), and strongly defended here by Podos alone. All of this is achieved by what Gelman calls "researcher degrees of freedom," and "p-hacking," and more.

Swamp Sparrows

I'm not going to go here in any kind of detail. I'm going to do my best to stick with chipping sparrows as a proxy of all that is ill in Podos' response [but I learn later than I'm a miserable failure at achieving that goal].

Kroodsma presents an original data set on swamp sparrows that suggests, as with chipping sparrows, that young birds copy song models with high accuracy. This is illustrated in apparent regional variations in song structure, and in how songs tend to cluster together to type in multivariate space (Kroodsma, 2017, his Figures 6e8). Thus, as with chipping sparrows, Kroodsma questions whether young male swamp sparrows could ever tailor songs to their own individual proficiency, tempting, but I'm going to go on. given the apparent premium on song learning accuracy.

Please cite this article in press as: Podos, J., Birdsong performance studies: reports of their death have been greatly exaggerated, Animal Behaviour (2016), http://dx.doi.org/10.1016/j.anbehav.2016.12.010

Once again, resolution of the issue requires attention to the song learning process itself. Here Kroodsma turns his attention to two laboratory studies, Podos, Peters, and Nowicki (2004) and Lahti, Moseley, and Podos (2011). The former study, which followed directly from Podos (1996), asked how male swamp sparrows respond during development to the challenge of copying songs that exceed their own vocal performance capacities. The conceptual basis for these studies was as follows. In nature, birds typically copy tutor songs that they should be able to reproduce faithfully. Yet over evolutionary time, song types might be selected for increased performance levels, which might push those songs to certain ends including faster trill rates. To simulate this scenario, Podos et al. (2004) trained young males with song models that were experimentally altered towards higher trill rates. The young males proved unable to reproduce these songs faithfully, and instead generated motor solutions that involved either reproducing song models at slower rates, maintaining the faster model trill rates yet omitting notes, or inserting pauses between multisyllable segments (Podos, 1996; Podos et al., 2004). The design of Lahti et al. (2011) followed the prior studies but also included models whose trill rates were decreased rather than increased. A primary finding of Lahti et al. (2011) was that when copying slow-trill models, birds adjusted the trill rates of these models upwards, towards more natural trill rates, thus enhancing trill performance at the expense of trill rate copying accuracy. All three of these laboratory studies illustrate that swamp sparrows, like chipping sparrows, are highly flexible in their sensorimotor learning and are able to develop songs very different in structure to the models from which they were copied.

Rather than conceding this point, Kroodsma's critique focuses on two other points. First, he argues that birds in these studies might have modified song structure during development not to maximize performance outcomes, but rather to develop normal, species-typical songs. The main outcome of Lahti et al. (2011), that birds reproduced slow models with enhanced trill rates, is indeed consistent with both performance and 'normalization' hypotheses. However, results from copies of fast learning models, in all three studies, fit more squarely in line with performance hypotheses. While some fast models were reproduced at slower trill rates, the structure and learning trajectories of other copies revealed efforts by birds to retain the faster model trill rates, which again was only achieved at the expense of other features including the loss of notes and the alteration of standard song syntax. These outcomes, and the asymmetric outcomes of the fast and slow model studies, cannot be explained by a normalization hypothesis. A performance constraint hypothesis (my H1) remains the best hypothesis standing.

13) Deception: Failure to disclose simple alternative explanations (again)

OK, given that the focus is now on Lahti et al. (I was going to spare Lahti, because I have come to respect him as a solid biologist/ornithologist, willing and eager to exchange ideas about how science is done, and I have come to believe that this introduction to Lahti et al. was imposed on Lahti by Podos, though Lahti won't talk about it), let's look more carefully at what I regard as a highly deceptive introduction. Here is the tell-tale quote from the introduction, with my bold-face words stressing the three possible outcomes that the authors present:

> ***If birds emphasize imitative accuracy over performance**, we would predict that (1) birds would learn and reproduce slowed models with high accuracy. On the other hand, **if birds emphasize vocal performance at the expense of accuracy**, we would predict that (2) birds would memorize slowed models yet reproduce them at higher trill rates. Another possible outcome is that experimental manipulations would render songs inappropriate as models, so that (3) **birds would ignore song models with diminished trill rates**. Distinguishing among these outcomes allows us to draw inferences about the relative importance of learning accuracy and vocal performance in song development*

Note that nowhere among those possible outcomes is 4) **the possibility that males will just strive to produce as normal a song as possible**, based on what they heard. That is the default outcome, the baseline that one might expect of a young bird who, in his genes, just might know what a good song should be. This fourth (rather boring) possible outcome is not mentioned. Why not? To me, not mentioning the obvious is highly deceptive; it's not how we do science (nor is it how we naturally navigate our way through life—see later). Now, only after I point out that fourth possible explanation, it is dismissed by Podos in a series of arguments that I am not going to study all that closely, because . . .

14) Blatant duplicity, up front, at the top, in the Title: "A Tradeoff Between Performance and Accuracy in Bird Song Learning"

Concealing explanation #4 allows the authors to generate the above title, which sounds pretty sexy, but is highly misleading.

Consider the following example: The trill rate of a normal song is reduced to 65% by adding silent intervals between the notes, thus creating a highly abnormal song. A young swamp sparrow, hearing this song, (instinctively) removes the silent intervals and learns to sing an entirely normal song, at 100% of the normal trill rate.

Here is the "main finding" of the authors: "*Our main finding is that birds elevated the trill rates of low-performance songs, but at the expense of imitative accuracy. . . The elevation of trill rates of slowed models supports the hypothesis that birds calibrate learned vocal output to match their individual performance capabilities . . .*" (What the authors don't admit is that the "low-performance songs" are highly abnormal.)

In other words, of the three possible explanations that the authors give for the data, explanation #3 can be dismissed because the birds actually learned the songs. Explanation #2 can be dismissed because the birds didn't learn the slowed abnormal songs, thus showing that accuracy in learning was less important than singing a high performance song. The authors and readers are left with explanation #1, that birds "*calibrate learned vocal output to match their individual performance capabilities.*" There is no other explanation for these data. The data are *consistent with* and *support* the hypothesis of calibration and performance and blah blah blah.

15)To which I say THIS IS 100% BULLSHIT (defined below).

> ***On Bullshit** (2005), by philosopher Harry G. Frankfurt, ... Frankfurt determines that bullshit is speech intended to persuade (a.k.a. rhetoric), without regard for truth. The liar cares about the truth and attempts to hide it; the bullshitter doesn't care if what they say is true or false, but rather only cares whether or not their listener is persuaded . . . (Wikipedia)*

Back in 2004, the last time Podos would ever speak to me or communicate with

> *"Science is the search for truth regardless of how good the story is, whereas 'marketing or advertising' is the search for a good story regardless of the truth."* – Donald Kroodsma to Jeffrey Podos, UMass Biology, 4 October 2004

In my opinion, Podos was marketing flimsy stories back then, and only after a 10-year hiatus ("retirement") would I return and discover how much marketing he had managed in the meantime.

Now, only when the duplicity of the authors has been exposed does Podos try to explain away explanation 4. But that's too late, because . . .

16) Authors who fail to reveal obvious alternative explanations in publications have no credibility when trying to explain them away later

Frankly, I have no confidence in repeated (as in publication after publication), after-the-fact denials of something that should have been up front to begin with. Also, I inevitably ask what else is being hidden from readers that they should know about? What other liberties (i.e., "researcher degrees of freedom"—Gelman) have the authors taken to develop their story?

In other words, if authors deceive in their original literature, why should their cover-up of it be believed when they are exposed?

Kroodsma's second point is that the focal papers did not demonstrate that a young swamp sparrow will 'adjust features of what he learns…to calibrate a normal, wild-type song to his own abilities…so that he can honestly broadcast his own individual quality' (Kroodsma, 2017, pp. e8ee12). However, the papers in question were agnostic to the question of individual variation in performance capacities. These papers made no mention of the concept of male quality, and made no attempt to characterize individual variation in any quality metric. Moreover, because the studies featured training models with manipulated trill rates, they could offer only limited direct insights into how normal songs are learned; that would take some other design. It seems unfair to critique papers for not answering questions they had not set out to answer. In any case, a possible follow-up experiment would be to somehow manipulate or quantify variation in male quality, and to ask whether and how such variation predicts the development of vocal performance features (Podos et al., 2009).

Kroodsma's citation of Podos et al. (2004) is also misleading in his suggestion that this was one of two papers that founded 'the performance hypothesis' (sensu Kroodsma, 2017). As noted above, Podos et al. (2004) made no mention of male quality (a key component of H1), nor did these authors include any reference whatsoever to song function (H2). In addition, in his critique of Lahti et al. (2011), Kroodsma presents three quotes that are all attributed to the wrong paper; the last of these quotes is presented as referring to biases in song production, whereas the actual quote referred to biases in song learning, a very different thing.

VOCAL PERFORMANCE AND SONG REPERTOIRES

Kroodsma's second main critique focuses on species with song repertoires, and might be summarized as follows. (1) The performance hypothesis posits that birds' songs reflect individual variation in their production capacities, with higher-quality singers able to learn and produce songs with faster trill rates and/or wider frequency bandwidths (my H1). (2) Yet, song types within individuals' repertoires tend to diverge widely in values for trill rate and frequency bandwidth, with some song types achieving only low values in these performance metrics. (3) Moreover, trill rates and frequency bandwidths of shared song types are highly consistent across birds, showing less structural variation within types than between types (within individuals). (4) How then could trill rate and frequency bandwidth values in repertoire species provide reliable indicators of singer quality? And relatedly, why would high-quality singers ever learn low-performance song types?

Kroodsma supports points (2) and (3) above with his original data set on swamp sparrows. As with other metrics of performance, within-individual variation in trill rates and frequency bandwidths is seen to exceed across-bird, within-type variation. This demonstration is not surprising, given the structural diversity of song types within populations. Indeed, similar results have been reported elsewhere (Cardoso, Atwell, Ketterson, & Price, 2009; Podos et al., 2016).

17) This isn't worth indexing:

Can't resist a comment here, unfortunately. In a draft of my Forum article, available to Podos December 2014, I point out three matters of some importance:

1) Cardoso et al. have interesting data on "song performance" that refute Podos' deviation hypothesis

2) Here are some more data on chipping sparrows and swamp sparrows that agree with Cardoso

3) Interesting how Cardoso has been excoriated by Podos in Zollinger, Podos et al. (2012), and never cited in a positive voice. Cardoso's work that effectively refutes the deviation hypothesis is never mentioned, despite numerous opportunities in publications by Podos.

With my Forum article in hand making the above three points, some months later, during 2015, Podos and his students (2016) submit their paper that repeatedly lauds Cardoso for his work and, furthermore, tells that my Forum points 2 and 3 above aren't all that new because "similar results have been reported elsewhere (. . . Podos et al., 2016)." I guess that's progress.

The open point of debate is thus as follows: acknowledging points (2) and (3), are there viable answers to the questions raised in point (4), or are the questions merely rhetorical? Kroodsma adopts the latter position, stating that, in repertoire species, '[p]erformance measures simply cannot be used…to assess the relative quality of a singer' and '[t]he data provide no support for the feasibility of the performance hypothesis' (Kroodsma, 2017, page e7). Kroodsma's point might seem reasonable at first glance. After all, if an individual bird's songs diverge widely in performance, how can those songs indicate singer quality? Yet there are (at least) two scenarios, reasonable yet mostly untested, that might provide answers to the questions raised in point (4).

The first scenario builds on Logue and Forstmeier (2008), a key paper in the field that Kroodsma failed to even acknowledge. Logue and Forstmeier (2008) hypothesized that, in repertoire species, listeners evaluating singers' vocal performances should be selected to hone in on song types shared by neighbours, such as those used during song type matching. This is because perceptual assessments of singer attributes should be relatively feasible to conduct across exemplars of a common type. By contrast, it should be more challenging to compare performances of song exemplars across types (i.e. for unshared types), because such exemplars will vary not just in performance but also in other baseline structural properties, with the latter potentially obscuring detection of the former (for a review of this and related principles, see Bateson & Healy, 2005). As Logue and Forstmeier (2008) noted, receiver bias towards comparing shared song types should in turn impose selection on birds to produce high-performance versions of shared types.

Logue and Forstmeier's (2008) paper focused on song type matching during territorial interactions. Yet I would argue that

Please cite this article in press as: Podos, J., Birdsong performance studies: reports of their death have been greatly exaggerated, Animal Behaviour (2016), http://dx.doi.org/10.1016/j.anbehav.2016.12.010

their model for song perception and assessment could apply readily

and more generally to additional circumstances, such as in species that do not engage in song type matching, or when birds are evaluating the songs of solo singers. If listeners retain perceptual and memory-based templates of standard performance levels for a population's shared song types, then listeners should be able to detect deviations from those standards. In other words, perhaps birds can detect the performance of birds' singing relative to type, as opposed to along some absolute scale. This scenario would allow between-type variation to exceed within-type variation, while maintaining the possibility that song conveys individual differences in vocal performance.

18) Is there any way to falsify anything?

Is there any way to falsify anything, or can scenarios and hypotheses always be proposed to perpetuate the pursuit of this line of thought and the urgent need for more grant money to do the work? Andrew Gelman's words come to mind, as this deviation/performance literature illustrates well . . .

> *" . . . the paradigm of the open-ended theory, of publication in top journals and promotion in the popular . . . press, based on 'p less than .05' results obtained using abundant researcher degrees of freedom. It's the paradigm of the theory that . . . is 'more vampirical than empirical—unable to be killed by mere data'."*

At this point it is worth emphasizing that different song types will likely differ in performance requirements beyond that captured by trill rate and frequency bandwidth. Other song features that reflect singer performance include the number of notes per syllable, the magnitude of amplitude fluctuations within and across notes, and the evenness of frequency transitions among notes in sequence. Variation in the latter factor was illustrated by Podos et al. (2009, their Figure 1; see also Podos et al., 2016), for two hypothetical song types with identical trill rates and frequency bandwidths, yet for which one should be harder to produce than the other. Similarly, song types that differ on trill rate and frequency bandwidth plots might have similar or even identical performance requirements. We should thus be cautious, when comparing song types, in using trill rate by frequency bandwidth values as absolute metrics of performance.

The second scenario directly challenges Kroodsma's declaration that 'an important condition for honesty and reliability is that males consistently use songs within a relatively narrow range of performance abilities' (Kroodsma, 2017, page e11). To the contrary, there are no good biological reasons to suppose that birds must produce all of their songs at maximal performance capacities, or that all song types must be selected to provide reliable indices of singer quality. Taking a step back, displays across the animal kingdom can be complex and multifaceted, and different display components might be shaped by distinct selection pressures. For example, some display components might be optimized for transmission, or for directing receiver attention, or to aid species recognition, or to denote distinct aspects of individual quality. The diversity of functions among distinct display components has been widely documented (e.g. Doucet & Montgomerie, 2003; Gibson & Uetz, 2008; Patricelli & Krakauer, 2009), including in recent work by Barske and collaborators on golden-collared manakins, Manacus vitellinus (Barske et al., 2014; Barske, Schlinger, & Fusani, 2015; Barske, Schlinger, Wikelski, & Fusani, 2011). This research team has shown that mating success of male golden-collared manakins is predicted by a limited set of display components, the vigour of 'wingsnaps' and 'rollsnaps', that likely push the boundaries of birds' mechanical and metabolic performance capacities. Yet there are other components to these bird's displays that are lower performance and that do not predict female choice, yet which presumably still serve other functions.

Returning to birdsong repertoires: at least two potential functions for low-performance song types are suggested by recent papers not cited by Kroodsma. First, low-performance songs could complement high-performance songs in dynamic, time-varying interactions, as animals escalate or de-escalate their signals of aggressive intent (Hof & Podos, 2013; de Kort, Eldermire, Cramer, & Vehrencamp, 2009; see also Searcy & Beecher, 2009). Recent studies of escalation have focused on song type sharing and use of low-amplitude songs (Searcy & Beecher, 2009); by contrast, little is known about possible roles of between-type performance variation in escalation. It could be that birds begin interactions using low-performance songs, and then switch to higher-performance songs as interactions escalate. Second, low-performance songs could be produced before or after high-performance songs in ways that enhance perception of the latter through a contrast effect. Lyons, Beaulieu, and Sockman (2014; see also Caro, Sewall, Salvante, & Sockman, 2010) illustrated this interesting possibility in an experimental study of female preferences for songs in Lincoln's sparrows, Melospiza lincolnii. This research team found that females' responses to species-typical songs covaried with performance levels of songs they had heard previously. More specifically, females previously exposed to low-performance songs responded more favourably to species-typical songs than did females previously exposed to high-performance songs.

We have much to learn regarding how birds' use of song types in nature might covary with inter-type performance variations. The scenarios I presented are largely untested, and it is not hard to envision experimental designs that could be applied towards these ends. For the time being, however, Kroodsma's blanket dismissal of this area of research ('the hypothesis becomes biologically implausible, if not impossible'; Kroodsma, 2017, page e1) is surely premature.

VOCAL PERFORMANCE AND SONG FUNCTION

Kroodsma's third main critique focuses on song function, calling into question reports that birds discriminate song performance variations and modulate their behaviour accordingly (my H2 above). His comments on sparrows focus on four papers: Ballentine, Hyman, and Nowicki (2004), DuBois, Nowicki, and Searcy (2011), Goodwin and Podos (2014), and Moseley, Lahti, and Podos (2013). For discussion of other species I defer to the accompanying papers by Vehrencamp, de Kort, and Illes (2017, in this issue) and Cardoso (2017, in this issue).

The first of the sparrow papers, Ballentine et al. (2004), gauged female swamp sparrows' responses to naturally high- and low-performance song variations. Their experimental design anticipated Logue and Forstmeier (2008) by presenting individual females with variations within type. Thus, the authors avoided the conundrum of having to interpret response differences to stimulus pairs that would have varied in both performance and type identity. The validity of the main result from Ballentine et al. (2004), stronger responses to higher-performance songs, stands firm; what Kroodsma instead offers are alternative hypotheses that might also explain the outcome. I regard two of these as credible and worth further attention. First, Kroodsma notes that perhaps Ballentine et al.'s (2004) low-performance stimuli scored low not just in trill rate and frequency bandwidth but also in note consistency, which might have happened had the low-performance stimuli been recorded from yearlings. Thus, perhaps females were differentiating songs not on the basis of performance but instead on the basis of note consistency. Second, perhaps songs that were rated as lower performance were recorded at greater distances and thus contained more reverberation, which as a correlated trait could be the trigger for low responses by females.

Please cite this article in press as: Podos, J., Birdsong performance studies: reports of their death have been greatly exaggerated, Animal Behaviour (2016), http://dx.doi.org/10.1016/j.anbehav.2016.12.010

The same critiques could also be applied to the first experiment of DuBois et al. (2011), who presented Ballentine et al.'s (2004) stimuli to territorial males, thus allowing direct comparison of the female and male responses. The other studies under scrutiny, including the second experiment of DuBois et al. (2011; see also Caro et al., 2010; Illes, Hall, & Vehrencamp, 2006; Lyons et al., 2014), accounted for these alternative hypotheses by employing an experimental design in which low- and high- performance stimuli were constructed artificially, by increasing or decreasing trill rates of songs recorded in the wild. This was achieved by decreasing or expanding internote intervals. As such, all low- and high-performance pairs in these papers were matched with identical degrees of note consistency and reverberation, thus eliminating

Please cite this article in press as: Podos, J., Birdsong performance studies: reports of their death have been greatly exaggerated, Animal Behaviour (2016), http://dx.doi.org/10.1016/j.anbehav.2016.12.010

Kroodsma's alternative hypotheses. This positive aspect of the manipulation-based design received no mention by Kroodsma.

Kroodsma's critique of Goodwin and Podos (2014) here builds on another point: song stimuli with altered trill rates, as included in the Goodwin and Podos (2014) design, might elicit diminished responses during playback not because they express reduced performance, but because they are structurally abnormal. Yet the key feature of Goodwin and Podos's (2014) design is that birds were presented with test songs that had been either slowed down or sped up by the same percentage. No song stimuli were presented at their natural trill rates. Strong responses to the atypically fast songs, compared to weak responses to the atypically slow songs (that is, asymmetry in response strength to the two treatment conditions) was the basis for Goodwin and Podos's (2014) interpretation that high-performance songs elicit stronger aggressive responses. In other words, the atypical nature of altered trill rates was controlled for by the study's matched stimulus design, with stimulus pairs matched for abnormality, and stronger responses to higher-performance songs were demonstrated with clarity (another point not acknowledged by Kroodsma). That birds still gave strong responses to high-performance songs in spite of their atypical trill rates, as in Draganoiu, Nagle, and Kreutzer (2002), suggests that the functional salience of performance-related vocal traits is actually underestimated by these tests. The same point applies to the second experiment of DuBois et al. (2011), and to Moseley et al. (2013).

19) Manipulated songs are highly abnormal songs, but is that ever admitted or discussed?

"by the same percentage"? So it is assumed that songs at 75% and 125% are equally abnormal to the birds, so the experimental design is therefore perfectly balanced with abnormal stimuli? What if the birds don't assess abnormality on a percentage scale?

"stronger responses to higher-performance songs were demonstrated with clarity"—I am not sure what remains of clarity, now that this entire study apparently clings to one statistical test, among many, that reaches a little below 0.05 (Goodwin and Podos 2015).

Do any of these studies ***openly*** address the possibility that the manipulated songs they use might be abnormal and the birds might therefore simply respond less to them? I don't think so. Again, it is failing to be upfront about simple alternative explanations that is so astonishing in these studies. And when the alternative explanations are raised, they are always dismissed. As I say above, when such obvious and simple alternative explanations are not dealt with in an open fashion in an original publication, authors have already lost the credibility they need to effectively address the issue later, especially when all previously undisclosed alternative explanations can always be dismissed in one way or another.

Another Gelman quote seems highly applicable:

> *"the real 'conclusion of the paper' doesn't depend on any of its details—all that matters is that there's something, somewhere, that has p less than .05, because that's enough to make publishable, promotable claims about 'the pervasiveness and persistence of . . .' whatever . . . they want to publish that day. When the authors protest that none of the errors really matter, it makes you realize that, in these projects, the data hardly matter at all."*

Another way Kroodsma aims to bolster his case against these latter papers is by illustrating hypothetical examples of how songs with modified trill rates are rendered atypical. In his first such example for chipping sparrows, Kroodsma imagines a trill with a natural trill rate of 28 Hz being reduced to 7 Hz. That kind of song would indeed sound highly abnormal, with the new rate a mere 25% of the natural trill rate.

How about 35%? Would that be highly abnormal? I think so. From Moseley, Lahti, & Podos (2013): "Songs were slowed to 35–80% of their original trill rates . . ." You can do a lot of damage to a song by following this guideline (from Goodwin and Podos 2014): "We created stimuli by increasing or decreasing trill rate while ensuring the song was within the observed population range." In other words, manipulated songs are considered ok if their trill rate (nothing else need be considered) falls within the 'observed population range,' no matter how strange or abnormal the songs might become during the manipulations.

As far as we know, no published study on vocal performance has used that severe of a manipulation. By contrast, in the actual study on chipping sparrows (Goodwin & Podos, 2014), low-performance trills had trill rates that averaged ~70% of the corresponding song's natural trill rate. This is in the ballpark only of the least severe manipulation in Kroodsma's additional hypothetical examples, in which 28 Hz is dropped to 21 Hz (75%). In the studies we have performed, the distinctions between most of the low- and high-performance versions of given song types have been quite subtle, at least to our ears. A parallel exaggeration about the severity of stimulus manipulations was applied in Kroodsma's critique of Moseley et al. (2013); Kroodsma's Figure 10 illustrates a hypothetical manipulation that matches only the single most severe manipulation employed by Moseley et al. (2013).

The point being, what? Here's my point: If that "most severe manipulation," along with other "less severe manipulations," was included in the methods, the results are seriously whacko.

Yes, pretty severe. Again, in Moseley et al. no mention is made of the possible abnormality of the songs, an example of yet another paper failing to mention even the simplest of alternative explanations for their results.. Here's what I wrote in my original Forum article, addressing another confounding variable as well (stimulus quantity):

> For Moseley et al., "*The confounding variables of stimulus quantity and abnormality do not seem to be of concern to the authors when they conclude the following: 'territorial male swamp sparrows responded significantly less strongly to low-performance than to control- performance playback stimuli, consistent with our hypothesis that receivers should attribute limited threat to low-performance songs' (page 4).*"

The data are also ***consistent with*** the "hypothesis" that birds don't respond to highly abnormal songs, or respond less to shorter than to longer songs, but no mention is made of those alternative, uninteresting, unpublishable explanations. This is all story-telling at its best.

Finally, it is worth noting that birds' responses to the most highly manipulated songs, even at the lowest trill rates, ten- ded to be fairly strong and often at baseline levels for major response features including subject flight and song rates (see supplemental material in Moseley et al., 2013).

In addition to being rendered 'abnormal', Kroodsma notes that song stimuli with manipulated trill rates will be further compromised because the manipulation necessarily changes the 'quantity' of the stimulus (this comment is offered with respect to Moseley et al. (2013), but would also apply to the other trill manipulation studies). It is certainly true that manipulated songs would have a greater 'quantity' of song per unit time. Yet Kroodsma's suggestion of trying to control for quantity would introduce additional changes

to stimuli that would hamper comparisons across treatments. For example, maintaining the total song quantity for songs with elevated trill rates would render those songs shorter in overall duration. Simply stated, there is often no perfect experimental design (Wiley, 2003). In defence of our chosen design, I AM REALLY TIRED OF THIS NONSENSE. All I am asking for is honesty in dealing with the possible shortcomings of a study. No deception. Tell us what the alternative explanations are. There's no need to defend in retrospect if authors are upfront and honest about what they have done to begin with. And if authors aren't honest up front, there's no reason to expect them to be honest when pressed. This is no doubt the central theme in my critique. Avoid the deception, dishonesty, half-lies—tell the **WHOLE TRUTH. Scrap the marketing and self-promotion, scrap the stories at the expense of science; do science.**. while our artificial stimuli varied not just in trill rate but also in quantity, quantity differences would also distinguish natural high- versus low-performance variants of the same song type. In other words, our approach to experimental manipulation of stimuli matches how natural variants of the same song types differ in nature. What? Bullpoop! False! Not true at all. That said, it would of course be useful to apply alternative playback designs to test further the functional salience of performance- related song variations.

LOOSE ENDS

Kroodsma's main critiques of Goodwin and Podos (2014) were four-fold. I have already addressed two (see above). The third critique, that interactions among males are competitive and not cooperative, seems misguided because it references behaviours that have only been seen to occur pre-dawn. Playback trials and observations of coalitions were conducted post-dawn, and none of the types of competitive interactions that Kroodsma references were observed during coalition formation. I will note that the active pre-dawn behaviour of male chipping sparrows does imply that these birds are especially tuned to their neighbours and to their neighbours' songs, as demonstrated in a different way by Goodwin and Podos (2014).

20) Failure to mention undermining facts builds a good story but is just plain deceptive

Again, to me the original paper is just plain deceptive. **Nothing** is mentioned in that original paper about the competitive pre-dawn gatherings that had been described previously for male chipping sparrows in the study populations of Goodwin and Podos. Now, after this rather glaring omission is pointed out, we're asked to believe that those competitive gatherings aren't relevant and weren't mentioned because they didn't occur during the day? I don't buy it one bit. The published story isn't nearly as good if the reader realizes that there's more going on when male chipping sparrows gather. And, one might ask, just how did the authors recognize a coalition, and how does one distinguish it from what one sees during the pre-dawn? ***At the very least***, readers deserve to know that the authors have somehow distinguished the pre-dawn gatherings from the daytime gatherings.

1) Failing to mention in the original paper anything about how a chipping sparrow learns its songs (discussed earlier), and 2) failing to mention in the original paper anything about male-male gatherings in other contexts that just might undermine the assumption of coalitions, are a disservice to readers. Readers expect and deserve to be educated, not deceived; information withheld to generate a better story is deception, plain and simple. I am not sure how to distinguish it from fraud. It's the kind of deception that won Goodwin and Podos (2014) a best student paper at the ornithological meetings in 2014.

The fourth critique, that Goodwin and Podos (2014) ran tests that were undisclosed, is simply incorrect. There were no undisclosed tests associated with Goodwin and Podos (2014), and Kroodsma's admonitions on this point are thus moot. (For the record: Kroodsma's critique here perhaps built upon Akçay and Beecher (2015), who asserted that Goodwin and Podos (2014) ran an undisclosed test on vocal deviation (with no mention of frequency bandwidth, as Kroodsma has added). Akçay and Beecher's assertion, however, was based on a personal communication from S. Goodwin that was misconstrued. To complicate matters, Goodwin and Podos (2015) committed an error of omission by failing to negate Akçay and Beecher's (2015) errant assertion. Let the present statement correct the record).

21) The above is a pants-on-fire paragraph.

First, the second author (Podos) is claiming that the first author didn't know what analyses she (Goodwin) did when she communicated with others about her work. It's pretty difficult to "misconstrue" what Goodwin conveyed, not only what she wrote to Akçay and Beecher but also in her discussions with me at the ornithological meetings during 2014. It hasn't helped that the authors since July 2014 would respond to no one about their work, threatened me with criminal harassment charges for asking, and instead went into a circle-the-wagons mode and deep secrecy—so nobody could inquire about how the work was actually done.

Second, the authors had a chance to correct this glaring misconception in their 2015 response to Akçay and Beecher, but somehow overlooked correcting it at that time.

Third, just think about it: Look again at that figure 4 I've reproduced above. The only reason this whole enterprise is interesting is if there's a tradeoff between trill rate and frequency bandwidth that somehow conveys something about how difficult a song is to produce and something therefore about the quality of the singer. That's why that figure was presented by Goodwin and Podos (2014); they didn't present a figure of how just the trill rate varies among birds. Now Podos would have us believe that the scatterplot isn't interesting at all? And he had no interest in checking whether there was any chance that males interacted with each other based on this tradeoff of trill rate and frequency bandwidth? In fact, frequency bandwidth wasn't even worth checking by itself, apparently, at least according to the claim by Podos now. The only statistics that were done in exploring the data were on trill rate?? In spite of the rationale for doing the work in the first place, and in spite of all the promotion of the tradeoff illustrated in that scatterplot figure over the years and years, I find it a challenge to believe that there was no interest in doing any statistics to check out the effect of frequency bandwidth or its interaction with trill rate.

And, in the end, none of the above really matters, as it is all a distraction from the one, lonely statistical test that remains for the authors to cling to for their story.

Kroodsma argues throughout that prior support for 'the performance hypothesis' (sensu Kroodsma, 2017) has been pervasive and uncritical. To reinforce his argument, Kroodsma deploys three quotes. Goodwin and Podos (2015) are quoted as stating that the performance hypothesis 'has been adopted widely in tests of song function' (page e1); Wilson, Bitton, Podos, and Mennill (2014) are quoted as naming the performance hypothesis 'a premiere illustration of how performance constraints shape the evolution of mating displays' (page e1); and Podos et al. (2009) are quoted as offering the following uncritical support for the performance hypothesis: 'Emerging descriptive and experimental evidence thus suggests [sic] that vocal performance varies among individuals, and suggests that singers who maximize vocal performance gain advantages in song function and ultimately in reproductive success' (page e11).

22) Do I really mis-cite these three quotes for my "apparent advantage"? Here's the evidence

22a) Quote #1. Hogwash!

These quotes are, however, cited incorrectly and to Kroodsma's apparent advantage. The original quote from Goodwin and Podos (2015, page 170) actually refers to a metric of vocal performance, vocal deviation, not to 'the performance hypothesis'. A technical comment about the method used to measure vocal performance was thus repackaged as evidence for uncritical thinking. Huh? Is that what I did? My beef specifically is with deviation. I think this goes back to the word game that Podos set up in the introduction to this paper; I glossed over it there, but assumed he was up to something that would reveal itself later, and here it is.

Reluctantly, I look up the way I used that quote and offer it here, in its entirety (emphases added here):

> *"The interesting hypothesis is that how close a song plots to the upper bound [i.e.,* ***deviation****] might reveal the difficulty of producing that song, so that songs near the upper bound honestly reveal a high-quality singer; both prospective mates and competing males might then use those high-performance songs to detect high-quality singers.* ***This hypothesis has 'been adopted widely in tests of song function'*** *(Goodwin & Podos, 2015, page 1), is touted 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, page 214), . . ."*

I can't believe I have to do this, but there it is. I clearly spell out that I'm writing about the deviation of a song from the upper bound, and how those songs near the upper bound (with low deviation) are considered high-performance songs, and how that hypothesis [low deviation = high performance] has been adopted widely, not to mention here yet again how ridiculous all that has become.

I don't get it. What's going on here? This is simple English usage. How can what I have written be so misconstrued by Podos? Is it an honest mistake on his part? Am I the only one who reads this and think it's pretty straightforward?

22b) Quote #2. More Hogwash!

Similarly, the Wilson et al. (2014, page 214) quote referred to another subject:

'Studies of trilled vocalizations provide a premiere illustration of how performance constraints shape the evolution of mating displays…' **Studies of trilled vocalizations encompass more than trill rate by frequency bandwidth scatterplots** (reviewed by Podos et al., 2009).

Bullshit! Here's the full quote from Wilson et al:

> *Studies of* **trilled vocalizations** *provide a premiere illustration of how performance constraints shape the evolution of mating displays.* ***In trill production****, vocal tract mechanics impose a* **tradeoff** *between* **syllable repetition rate** *and* **frequency bandwidth****, with the* **trade-off** *most pronounced at higher values of both parameters.*

The quote from Wilson, Bitton, Podos, and Mennill (2014) is about trilled vocalizations and why they're interesting, as clarified in the second sentence, because of the tradeoff between "syllable repetition rate" (i.e., trill rate) and "frequency bandwidth." For Podos to claim that these two sentences at the beginning of the abstract do not go together, and that the first sentence needs to be taken in isolation and out of context, is hogwash, pure and simple. What Podos has done with these two quotes is, well, there it is for anyone to see.

Yes, of course, "Studies of trilled vocalizations encompass more than trill rate by frequency bandwidth scatterplots"; everyone will agree with that statement, but that's not what is conveyed in those first two sentences of the Wilson et al. abstract. I do not begin to understand how anyone could claim otherwise.

Moreover, it is worth noting that the main point of Wilson et al. (2014) was to challenge, rather than to confirm, a standard statistical approach used to quantify vocal performance.

Great! Here's now to measure a meaningless value more reliably

Podos apparently asks for some credit with this apparent self-flagellation, as if he is being critical of his earlier work. I'm not too quick to give credit, as all of this is sheer folly: The basic point made by Podos here and in Wilson, Bitton, Podos, and Mennill is that earlier measurements of a meaningless value (upper regression bound and therefore "deviation") were not done so well, but now we have an improved method so that we can measure this meaningless value more reliably. That's great!

22c) Quote #3. Shame on me.

In the third quote, Kroodsma altered the first verb from 'indicates' to 'suggests'. At the surface this might be taken as a minor transcriptional error, yet the original wording had been chosen with care to contrast the

Please cite this article in press as: Podos, J., Birdsong performance studies: reports of their death have been greatly exaggerated, Animal Behaviour (2016), http://dx.doi.org/10.1016/j.anbehav.2016.12.010

strength of the two conclusions: available evidence indicates that vocal performance varies among individuals (we can measure phenotypic variation with confidence and thus quantify individual variation with rigour, H1), yet can only suggest that vocal variation has functional consequences (data about signal function are always harder to garner and interpret, with firm conclusions always more elusive, H2). Moreover, the original quote referred not only to trill rate and frequency bandwidth but to the field as a whole.

Here's the quote as I transcribed it:

> *'Emerging descriptive and experimental evidence thus suggests [sic] that vocal performance varies among individuals, and suggests that singers who maximize vocal performance gain advantages in song function and ultimately in reproductive success' (page e11).*

Well, you know, I'm not even going to double check this one. The words that jump out at me here are these: the "wording had been chosen with care." It's worth reflecting on those and what has been intentionally conveyed throughout Podos' response in this document. Is this really how Podos thinks, that I would alter a word to score points?

But I will say that "indicates" is far too strong a word for Podos to use in the original quote, given the lack of evidence. When I substituted words, maybe I thought I was fixing the wording for him.

Following up on the prior point: The quote from Podos et al. (2009) illustrates a central feature of the modern literature on vocal performance, which is that the relevant questions, tests and hypotheses are typically separated into two main components, phenotype and function (H1 and H2). The literature has in fact followed, very precisely, the recommendation of Marler and Hamilton (1966), as quoted (with another incorrect transcription) by Kroodsma (2017). His admonition here, that we do not provide clear separation of description and function, is thus perplexing.

More on this later.

Kroodsma's original data sets are both incomplete and nonquantitative, and should thus be regarded with caution. I offer two specific illustrations. First, Kroodsma (2017, page e2) asserts that 'a young male chipping sparrow learns rather precisely the song of his adult tutor, and especially the tutor's trill rate'. While this assertion might seem to be supported simply by looking at spectrograms, it would require evidence that young males indeed develop trill rates closer to those of their tutors than to other males in the population that sing the same trill type (e.g. the circles enclosing data points in Figure 5 of Kroodsma, 2017). That in turn would require a much broader sample and some sort of statistical test. In addition, even if such a test were offered, mere demonstration of acoustic similarity of neighbouring males would not itself provide evidence that the subject learned from that tutor. That would require a controlled test of learning, as in Liu and Kroodsma (1999). Finally, in contrast to Liu and Kroodsma (2006), Kroodsma's original data set could not differentiate birds by age, so there is no way of knowing whether the neighbouring birds in Kroodsma's data set that shared the same song type were actually tutor and tutee, as opposed to adults who happened to share the same song type.

I agree, to a point, but mostly this is all beside the point, small points made here designed to create doubts when the big picture is abundantly clear. The graphs illustrate how neighbors share songs, and the mechanism for that has been demonstrated by following hundreds of young birds during dispersal. The pairs of similar songs are like those that have been demonstrated for tutor-tutored pairs, but when one finds a pair of birds like this in nature the best one can say is that they are part of the same mini-cultural tradition (involving surprisingly few birds given that there are so many different mini-cultural traditions) that transmits the same song type from one generation to the next. . . . I'm not sure what the point is of Podos' paragraph, but it does effectively distract from more important issues.

The second illustration concerns Kroodsma's measures of frequency bandwidth as they vary with distance and equipment (his Figure 9). There appears to have been only one sample taken per distance per recording set-up, which makes it difficult to accept the validity of the measures presented. Random variation in sampling conditions might be sufficient to swamp out real patterns of interest. Consistent with this point, the reported increase in frequency bandwidth measures from 8 to 16 m defies expectations. In addition, there is no attempt to provide statistical confirmation of the patterns presented. Finally, measures of frequency bandwidth taken for the original digital file ('0' metres condition) are simply not comparable to measures derived from re-recorded samples, because the original digital file did not incorporate the influence of the playback system on song structures. No inferences at all can be gleaned from that particular comparison.

23) My graphs of frequency bandwidth over distance with different microphones could be all wrong, or not.

Or they could be all right, or there could be elements of wrongness or rightness, or whatever. They might illustrate some special circumstances under which I did my comparisons. I don't know, but I wouldn't be so quick to dismiss them if this area of research were my passion and I relied on frequency bandwidths to make my arguments (more on this below, for the effect of analysis bandwidths of measured frequency bandwidth).

A further note on this second example: Kroodsma asserts, without justification, that the reliability of frequency bandwidth measures is 'assumed wrongly'. On the contrary, published work has already made clear that measures of frequency and thus frequency bandwidth are distance and condition dependent (e.g. Naguib et al., 2008). Because of this, frequency bandwidth is indeed best evaluated at close range to singers, by both birds and researchers. This is not a special problem, however, for the matters at hand; all scientific data include noise, and what is important is whether that noise might bias the outcomes of interest. In analysing trills, there is no reason that I know of to think that distances from which natural recordings are made are biased along some performance gradient. Kroodsma's critique here thus does not undermine the validity of analyses such as those presented by Wilson et al. (2014).

DuBois, Nowicki, and Searcy (2009) presented territorial males with playback of two stimulus classes, heterospecific song and conspecific song, and observed that vocal performance was higher in response to the latter. The differences in vocal performance values were minor but occurred in the same direction consistently across birds. Kroodsma asks readers to dismiss the validity of this main result, the uptick in performance, on two grounds. First he questions whether the slight uptick in vocal performance could be functionally meaningful. DuBois and collaborators tested this very point empirically in a follow-up study (DuBois et al., 2011), yet it seems that Kroodsma somehow expected functional questions to already be resolved in the earlier report.

Please cite this article in press as: Podos, J., Birdsong performance studies: reports of their death have been greatly exaggerated, Animal Behaviour (2016), http://dx.doi.org/10.1016/j.anbehav.2016.12.010

24) Bullpoop! Here's what I expect in good science: an honest, open evaluation of data and the consideration of alternative explanations for those data,

all in an attempt to determine what birds actually do. That happens immediately, not delayed by two years. When the obvious is concealed, everything else about a study is questionable as well. The trill rate holding up, as is claimed in the next paragraph? How much credibility do the authors have at this point? On anything. Very little. For me, they have zero credibility. The authors claim offense? They have brought it on themselves, over and over and over again.

Where in DuBois et al. (2009, 2011) Kroodsma sees conceptual flaws, I see the applica- tion of a clear sequence of logic, in which natural patterns were first documented and their functional salience then tested. Second, Kroodsma suggests that birds responding to conspecific song stimuli might have been recorded at closer distances, which would bias the measures of frequency bandwidth from these birds' songs to higher values. This seems worth looking into. Yet I note that there would be no effect of distance on trill rate measures, which means the original trill result stands firm. Also, I would recommend against using Kroodsma's (2017) Figure 9 as a reference for how bandwidth changes with distance, given insufficiencies in this data set (see above).

Kroodsma also suggests that DuBois et al.'s (2009) title, '**Swamp sparrows modulate vocal performance in an aggressive context**' is misleading. His rationale here is that swamp sparrows show inherent variation in performance features that naturally span a similar range of variations shown in the experimental protocol. Yet baseline variations say nothing about DuBois et al.'s (2009) actual finding, which was that a statistically significant majority of birds sang at higher performance levels during conspecific playback versus during heterospecific playback. DuBois et al.'s (2009) title is in fact **exceptionally precise**.

25) Podos defends half-truths, which are half-lies, which are deliberate attempts to deceive

Here, then, is another exceptionally precise statement:

"Swamp Sparrows modulate vocal performance in **nonaggressive** contexts."

No mention is made of this precise statement in the article by DuBois et al.

Oh, they have statistics to back up their title? When I looked at those statistics, I wondered immediately why they were done the way they were. It seemed that they were done to conceal what individual males had done. When there is so much deception in this literature, why shouldn't I believe that the deception also extends to the statistics that are done? I doubt very much that they are the whole truth. Well, I know they're not, because the authors ignore entirely the effects of distance on the recordings, and they fail to mention that obvious problem. Now Podos admits that possibility, and then falls back on another statistic to confirm the original conclusion, which is the routine in these studies, as if the data do not really matter, as the results and the conclusions cannot be shaken.

But note that when one (frequency bandwidth) of the two items (trill rate and frequency bandwidth) in the tradeoff hypothesis is lost, no longer is there reliable information about deviation and "song performance," so the original idea about "performance" in the context of deviation is lost.

And trill rate alone now tells about performance and the quality of a male? Other species can also slightly "compress" their song when excited, e.g.., sing a trill slightly faster, as during the dawn chorus, thus reflecting higher motivation, or so it has been thought. There's a simple, alternative explanation for increased trill rate: higher motivation. But does that increased trill rate really reflect "performance ability" and the quality of the male? I doubt it.

If swamp sparrows modulate their songs in all contexts, then to state that they do so in aggressive contexts without mentioning any other contexts is a **HALF-TRUTH, WHICH IS A HALF-LIE,** which is what so much of this deviation/performance literature is. The following quotes from the internet about "half truths" or "half lies" describe much of this literature:

> Merriam Webster: *a statement that is only partially true, . . . a statement that mingles truth and falsehood with deliberate intent to deceive*

> *A half truth is but an unfinished story. It's not truly a lie as no falsehoods were spun, however it's not the "full truth" as part of the story was left out. Thus why suspects being sworn into court must state they "swear to tell the truth, the WHOLE TRUTH, and nothing but the truth."*

> *Well no offence but I just believe in clear truth & a clear lie, if somethings is a "half" truth, then it's not truth at all.*

> *A half truth is partial truth, relayed with the other facet which completes it unmentioned, resulting in a message that paints a grossly inaccurate picture in the listeners mind that falls short of actuality.*

> *A half truth is a half lie. It's the same thing as a glass being half empty or half full. In any event it's still a deceptive statement. It just has a root in truth so it looks believable on the surface. For example, a politician says "No new taxes". Then when he gets elected he raises the existing taxes. His intent was to raise taxes all along, however he didn't tell that part.*

> *Halve truths? Are nothing more or anything less . . . Than half of a lie?!*

> *A half truth is only telling some of the story and not all of it to cover up something you're trying to hide.*

Not to belabor the point, as it is all rather simple how the authors of this literature convince readers of their stories. There is NO WHOLE TRUTH to be found in this literature.

26) Significant measurement errors cannot be dismissed so easily

Kroodsma also notes that the magni- tude of the treatment effect in DuBois et al. (2009) fell within the frequency resolution limits of the spectral analysis. This too does not alter the validity of the results, which showed a reliable and statistically significant uptick in performance in spite of analytic limits.

Here's the quote from my Forum article:

> *It should also be noted that Dubois et al. (2009) measured frequency at a resolution of 172 Hz, yet the frequency difference between neutral and aggressive contexts was reported as 91 Hz, about half the magnitude of the measurement error, thus rendering their frequency measurements inadequate.*

Please cite this article in press as: Podos, J., Birdsong performance studies: reports of their death have been greatly exaggerated, Animal Behaviour (2016), http://dx.doi.org/10.1016/j.anbehav.2016.12.010

So the resolution in the measurement is 172 Hz, and differences between two groups were 91 Hz, so as long as one got a statistically significant difference between the two groups, no matter how big the measurement error, it's ok? Is that really how statistics work and how we are supposed to use them?

Perhaps I should offer some information that, in the end, I chose to leave out of the Animal Behavior reply, for brevity, for simplicity, because the editors, referees, and I came to agree that I should make it short, as "less is more."

Here's the additional problem: The errors introduced in the analysis by a wide filter bandwidth (e.g., 172 Hz) are a function of the song type being analyzed (see the "web extra" on the opening page of my website devoted to these issues, copied below: http://donaldkroodsma.com/?page_id=1596). Using 5 different chipping sparrow songs, I found that "errors" (compared to a very narrow analysis bandwidth) ranged from near 0 up to 15% for a 172 Hz filter, as used by DuBois, Nowicki, and Searcy (2009). I suppose this also falls into the "all data have noise category," and as long as one got a statistically significant result, it's ok? I don't think so. Given errors up to 15% for just these five songs, I have to wonder how these kinds of measurement errors contributed to a statistically significant 1.8% increase in frequency bandwidth that was reported by Dubois, Nowicki, and Searcy.

Here is the "web-extra," copied below:

How wider analysis filter bandwidths lead to errors in measuring frequency bandwidth

for five different chipping sparrow songs

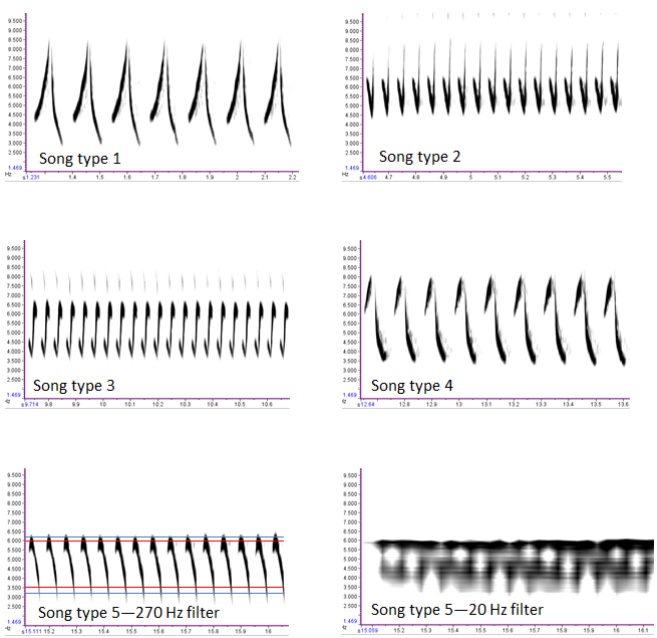

Figure 1. One-second excerpts from sonagrams of the five chipping sparrow song types used to explore how measures of frequency bandwidth and therefore "vocal deviation" are affected by the filter bandwidth used in software programs. Wide-band sonagrams (e.g., 270 Hz filter, as used on 5 of the 6 sonagrams) provide aesthetically pleasing sonagrams and accurate temporal measurements; narrow-band sonagrams (e.g., 20 Hz filter) are essential for measuring frequency accurately. For song type 5, the frequency bandwidth (at -24 db from max power) is shown for a 20 Hz filter bandwidth (red; 3578 to 6046 Hz) and a 270 Hz bandwidth (blue; 3270 to 6244 Hz; see also Figure 2).

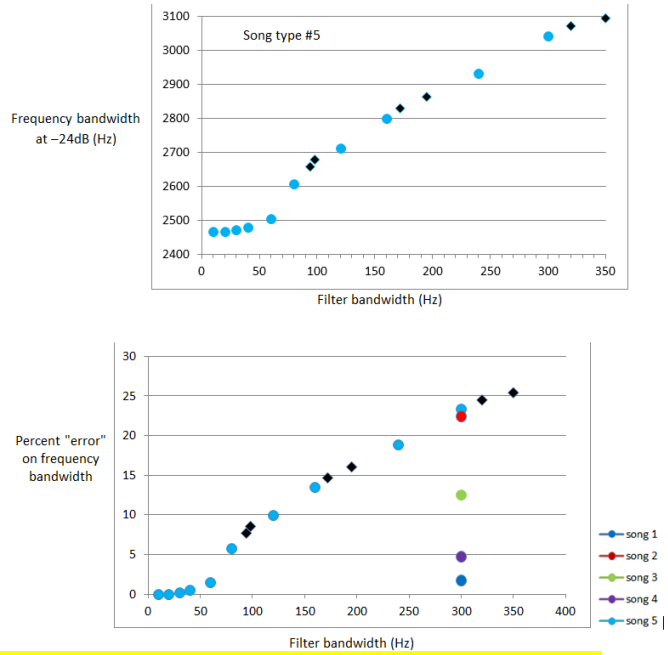

Figure 2. The filter bandwidth (i.e., frequency resolution) used in software programs strongly affects measures of frequency bandwidth. Best filter bandwidths for chipping sparrows are from 10-30 Hz (I used 20 Hz for analyses), after which resolution becomes increasingly less accurate. Primary curve is for song type #5 (Figure 1), with only the 300 Hz bandwidth plotted for songs 1-4; "percent error" is the difference in frequency bandwidth from that measured with a 20 Hz filter bandwidth. Frequency resolutions used by different authors range widely; in black diamonds are plotted some of the high extremes, at 350, 300, 195, 172, 98, and 94 Hz (Cramer & Price, 2007; Podos, 1997; Beebee, 2004; DuBois et al., 2009; Ballentine et al., 2004; Vehrencamp et al., 2013, respectively). How the filter bandwidth affects songs of other species awaits further description.

References

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

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

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, 122-127.

DuBois, A. L., Nowicki, S., & Searcy, W. A. (2009). Swamp sparrows modulate vocal performance in an aggressive context. Biology Letters, 5(2), 163e165. http://dx.doi.org/10.1098/rsbl.2008.0626.

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

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.

Kroodsma (2017, page e11) criticizes Ballentine (2009) for 'writing repeatedly that the data "support" the hypothesis that birds attend to performance ability'. However, Ballentine's (2009) study focused solely on the relationship between bird and song attributes (H1), with no commentary offered at all on song

perception (H2). Moreover, Kroodsma states that Dubois et al. (2011) used the same song playback used previously by Ballentine (2009). Yet, Ballentine (2009) did not run any playback trials.

27) Much belated applause for Cardoso et al.

Like Kroodsma, I applaud the research programme of Cardoso and collaborators on dark-eyed juncos, Junco hyemalis,

I would suggest there was no applause until I forced the issue in my Forum article, and once that was in Podos' hands, several months later (Podos et al. 2016), the first positive citation of Cardoso appeared in the Podos literature. Perhaps just a curious coincidence?

and other species. However, unlike Kroodsma, I do not regard our published commentary on methods for measuring frequency and amplitude (Zollinger, Podos, Nemeth, Goller, & Brumm, 2012) as an attempt to dismiss Cardoso's research programme 'on a technicality'. Rather, that commentary aimed to discuss, improve and standardize methods used in our field. It had exactly nothing to do with Cardoso's research on vocal performance and repertoires.

Please cite this article in press as: Podos, J., Birdsong performance studies: reports of their death have been greatly exaggerated, Animal Behaviour (2016), http://dx.doi.org/10.1016/j.anbehav.2016.12.010

28) Podos here articulates the essence of how to do science and makes this claim: "we used standard scientific practice in our approach . . ."

Throughout his critique, Kroodsma chides us for interpreting data as being 'consistent with' or providing 'support' for certain hypotheses. Yet, we used standard scientific practice in our approach, which involved articulating hypotheses, generating data and evaluating the fit of the data to the hypotheses. If data outcomes are consistent with a hypothesis, then the hypothesis stands, at least on a provisionary basis. Consistency of data with a hypothesis does not imply the hypothesis 'must therefore be true' (Kroodsma, 2017, page e7).

29) That's more hogwash: Let's explore that claim (see Zollinger, Podos, et al. for more hypocrisy)

Chide is an understatement. Closer to harangue. What is going on here when Podos makes that claim? I have some ideas . . . ~~I had written some pretty harsh, candid words here, and then chose to delete them.~~

What Podos writes is (more or less) the ideal of how to do science. We have multiple working hypotheses and try to falsify them until one is left standing. I love how Chamberlain (1965, The Method of Multiple Working Hypotheses, Science) articulated the problems if one is testing only one hypothesis, such as in the following quote (my emphases):

> *Conscientiously followed, the method of* ***the working hypothesis*** *is a marked improvement upon the method of* ***the ruling theory****; but it has its defects—defects which are perhaps best expressed by the* ***ease with which the hypothesis becomes a controlling idea****. To guard against this, the method of multiple working hypotheses is urged. It is directed against the radical defect of the two other methods; namely, the* ***partiality of intellectual parentage****.* ***The effort is to bring up into view every rational explanation of new phenomena, and to develop every tenable hypothesis respecting their cause and history****. The investigator thus becomes the parent of a family of hypotheses: and, by his parental relation to all, he is forbidden to fasten his affections unduly upon any one. In the nature of the case,* ***the danger that springs from affection*** *is counteracted, and therein is a radical difference between this method and thepercxent*
>
> *receding.* ***The investigator at the outset puts himself in cordial sympathy*** *and in parental relations (of adoption, if not of authorship)* ***with every hypothesis that is at all applicable to the case under investigation****. Having thus neutralized the partialities of his emotional nature, he proceeds with a certain natural and* ***enforced erectness of mental attitude to the investigation****, knowing well that some of his intellectual children will die before maturity, yet feeling that several of them may survive the results of final investigation, since it is often the outcome of inquiry that several causes are found to be involved instead of a single one.*

Podos and colleagues are not following "standard scientific practice," not by a long shot. I have laid out the problem repeatedly in my Forum article: It is the highly deceptive practice of finding data *consistent with* a favored idea, whether called a "ruling theory" or a favored hypothesis that has become a "controlling idea," and then declaring victory without also *honestly* disclosing to the reader (and perhaps to themselves?) what other explanations are consistent with the data. The other explanations typically aren't very interesting, and so it's no surprise that they aren't mentioned, especially if the author wants a good story. And good stories get published, lead to promotions, jobs, etc, as long as the authors are not caught and there's no cost involved. But that's all they are, good stories. They are not science.

I object enormously to the DECEPTIONS that occur paper after paper in the Podos-style literature.

30) The PODOS METHOD of ignoring alternative explanations fails spectacularly not only in science but also in everyday life.

Just imagine the research strategy of Podos trying to achieve something significant, like developing the atomic bomb, getting to the moon, or fighting an ebola outbreak. The failure would be catastrophic, because little, if anything, can be learned about the natural world using the research strategies that Podos defends. Progress in any scientific endeavor absolutely requires full disclosure and examination of everything, of all competing ideas and all competing explanations, or there's no progress. Zero. And furthermore, I claim that, because of the severe limitations of Podos' research, enormous damage has been done to this entire field of endeavor, because he and his collaborators been prolific and highly influential.

In everyday life, we constantly evaluate situations to solve problems. Using a rational approach, we do the best we can at identifying all possible explanations before settling on a solution, or a combination of solutions. That's what comes naturally. One has to learn to do the opposite; one has to work pretty hard to ignore alternative explanations, though many "scientists" seem to have come by that skill quite naturally.

I can come up with multiple examples of the kinds of things each of us does every day, in working through "multiple explanations/hypotheses" for something that we need to explain or accomplish. Here's not an everyday experience, but I think it makes the point:

> You have a loved one with an undiagnosed illness dying in the hospital. A rational person (you) would demand of the doctor all the (reasonable) possible explanations for the illness before coming to a decision about what to do next. The (ethical) doctor obliges, of course. With all information in hand, with all options on the table, a collective decision is made about the most likely explanation for the illness, and the course of action.
>
> Consider the alternative, what I think of as the Podos Method. You ask the doctor for one sexy explanation and ignore what are likely the real explanations. Patient dies in short order.

Why should the approach to science be any different from the natural way rational people know how to make decisions when something important is to be accomplished????

Ignoring alternative explanations for something in science renders the work worthless, and in real life ignoring those explanations can quickly become fatal.

Again, I just don't get it. If Podos has a hypothesis and finds data consistent with that hypothesis, then what? I suppose if the data don't fit the hypothesis, then it might be concluded that the data are wrong. If the data are consistent with the tested hypothesis, and if the alternative explanations for the data haven't been formally articulated as hypotheses, then he doesn't have to mention them because they're not hypotheses? The paper by **Lahti, Moseley, and Podos** (discussed above) is a prime example of this problem, as is **Goodwin and Podos (2014).** Another particularly egregious example is that by **Podos, Peters, and Nowicki (2004),** on "calibration," also critiqued in my Forum article.

Please cite this article in press as: Podos, J., Birdsong performance studies: reports of their death have been greatly exaggerated, Animal Behaviour (2016), http://dx.doi.org/10.1016/j.anbehav.2016.12.010

31) Cultural transmission of flawed research strategies, and the training of graduate students to think that this is science

I'll go right to the point. I know I'm not supposed to say any of this, because it is "ad hominem." Yet is it scientists who do the research, so inevitably it is the scientist who is being evaluated. The bold-faced citations at the end of the previous paragraph tell it all, as Podos learned this style from his advisor Nowicki, and Podos is transmitting it to another generation in his students.

The work by Podos et al. reveals a cultural transmission of research and publication techniques, across three generations, with "collaborators and former students . . . [and mentors all using] . . . similar research styles, favoring flexible hypotheses, proof-by-statistical-significance, and an unserious attitude toward criticism"; from Andrew Gelman blog).

An interesting historical note here: Some years ago, I was asked by Duke University to provide a letter of promotion for Steve Nowicki. I said I would write a letter, if they wanted it, but it would be a letter recommending demotion rather than promotion. The chairman of Duke's Biology Department then realized that he didn't actually need that letter from me. Only evidence consistent with the desired outcome was acceptable—sound familiar?

Is it any surprise that in the last 25+ years, I have never had the opportunity to review a manuscript or grant submitted by Nowicki (or Searcy for that matter)? And not since 2004 by Podos. I have always wondered what they say to journal editors and granting agencies about me, and perhaps about any others who might object to the style of publishing I am critiquing here.

Kroodsma also objects to our use of the word 'performance'. He actually professes to flinching upon our use of the word. It might be that Kroodsma's reaction stems from confusion about how we use the term. So to clarify: the term 'performance' draws specifically from the field of ecological morphology (Arnold, 1983; Garland & Losos, 1994; Irschick 2003; Wainwright, 1994), where it is used to account for complexities in relationships between morphology and behaviour, how that relationship can vary with context, and what that relationship means for evolution by natural or sexual selection (Byers, Hebets, & Podos, 2010). Many of nature's most fascinating behaviours involve animals pushing the limits of their physiological or mechanical performance, and these limits are often primary loci of natural selection (Irschick, Briffa, & Podos, 2015; Irschick, Meyers, Husak, & Le Galliard, 2008). A main goal in our papers in this realm has been to explore how the concept of performance translates into questions about birdsongs, specifically given demonstrated challenges involved in song production and development (e.g. H1). Of equal importance, the term as we use it does not aim to describe success (or failure) in communicative function (i.e. H2). In my view, using the term in this restricted way, as we have done, remains perfectly appropriate and does not prejudice subsequent assessments about whether and how performance variations are perceptually or functionally relevant.

32) Overuse of the word "performance" obfuscates and misleads . . .

Here is a brief section excerpted from my response that will appear in Animal Behavior, I which I explain my objection:

> *As I described in my Forum (Kroodsma, 2017), using the nonneutral 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 conclusions that songs with low vocal deviation are "better" and that birds with low vocal deviation therefore perform better and are 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 during December 2014.*
>
> *Try reading Podos et al. (2016) without the hidden implications of the word "performance," which occurs 139 times. 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 a good description, which is what all of this performance research would have benefitted from in the first place. 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.*

Finally, Kroodsma applies liberally a critique that our papers are tainted for not having used blind analyses of data. First, this is not fully accurate for some of the papers critiqued; where possible, we did employ blind analyses, even though these procedures were not explicitly stated in our methods sections. Second and more generally, in our field it is often difficult to collect or analyse data in ways that are completely treatment-blind. Experimental tests in our field are often intricate, the behavioural responses subtle, and the field sites remote. Behavioural ecology relies on the expertise of professionals to design the research, collect the data, and to generate neutral, objective evaluations of how data support or refute hypotheses of interest. In any case, claims of flaws in methods of data collection would ideally build on evidence rather than speculation.

CONCLUSION

In conclusion, I regard Kroodsma's essay as a low-performance enterprise that, by and large, fails to provide a reliable indicator of the quality and promise of this rich area of inquiry. More specifically: Kroodsma's presentation builds on a repeatedly inaccurate portrayal of published literature, on correspondingly questionable logic, on data sets that are incomplete and thus inconclusive, and on perplexing complaints about word usage. Even if one were to accept Kroodsma's critiques at face value (not my recommendation), there is no scientific basis for his outright rejection of performance hypotheses, for example in his declarations that 'song performance cannot be a reliable measure of male quality' (Kroodsma, 2017, page e11), or that 'there is no consistent, reliable information in the song performance measures that can be used to evaluate a singing male' (Kroodsma, 2017, abstract, page e1). Failing to support a hypothesis and rejecting a hypothesis outright are two very different things, with the latter requiring a stronger empirical and logical foundation. Kroodsma's case is also tainted in his implication that he is somehow able to evaluate how close song performance research comes to 'revealing truths about the natural world'. Nobody of course has direct access to the truth, which is why we do the science in the first place.

33) Given all of the deception, Podos' discussion of truth here rings hollow

This is not to say, of course, that our understanding of vocal performance is complete. To the contrary, as this exchange has highlighted, many open questions about vocal performance remain, especially regarding its interface with topics in vocal learning, repertoire development and song function. Here I will also acknowledge the validity of one of Kroodsma's underlying critiques: in spite of a growing body of studies on vocal performance, we still have limited direct evidence that high-quality singers in

Please cite this article in press as: Podos, J., Birdsong performance studies: reports of their death have been greatly exaggerated, Animal Behaviour (2016), http://dx.doi.org/10.1016/j.anbehav.2016.12.010

natural settings are able to develop songs with faster trills rates or wider frequency bandwidths, or that such differences matter in field interactions. Testing vocal performance hypotheses at this strictest level will be operationally challenging, and ideally would aim to (1) document variation among individuals, preferably in field settings, using some nonvocal metric of quality; (2) track social interactions and acoustic experiences for all young learners, so as to characterize the range of tutor songs each bird could presumably copy; (3) compare individuals' learning opportunities to vocal outcomes, to determine whether individual birds adjust song structure in accordance with their own quality; (4) and test whether performance-related vocal variations trigger differential responses from other birds. Individual studies have only achieved these components in isolation, and more integrative research programmes would be of great value. In the meantime, what can we say about available evidence for performance-based hypotheses in birdsong? There is no space here to review that evidence (I would update Podos et al., 2009), but I maintain that as a whole it is compelling, and enhanced by the fact that it has been garnered from diverse species, from across field and laboratory environments, and employing descriptive and experimental approaches. More work is required, yet for now the foundations of the hypothesis remain firmly intact.

34) An enormous waste of human resources and taxpayer money, and the long-term damage.

Self-evident. Nothing more needs to be said here, other than . . . Just look at all the effort put in by the authors to put these studies together, with nothing to show for it. And look at all of the federal grants (and other monies) used to do the work. And look at the dissertation improvement grants used to teach graduate students the art of Deception, on how not to do science. And think of the science that could have been done with monies invested elsewhere, with a different mindset. The long-term damage is to the scientific study of birdsong and sexual selection, and to the careers of those trained in this Art, and to science in general.

35) Those who defend this performance literature expose their own standards for science

As long as I'm addressing Podos' reply, I might as well here offer the most salient comment to the opening paragraph of Vehrencamp et al. (2017), considered eloquent by Podos, in which the performance literature I critique is strongly defended.

From Vehrencamp et al., their opening paragraph: "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 con- cludes 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). Those weaknesses include outright errors and misrepresentations, highly selective citation of the literature and convoluted logic (sensu Podos, 2017) . . ."

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 (Kroodsma 2017) and also in the detailed response to Podos in this document.

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.

35a) University of Massachusetts Amherst administrators and oversight committees also condone all that I expose here

. . . which makes one wonder what the overall standards are for science at the University, or what else is being hidden, or denied, or covered up. Details to be found at http://donaldkroodsma.com/?page_id=1596, section 5.

36) From Andrew Gelman's blog—p-hacking, research incumbency rules, researcher degrees of freedom, the garden of forking paths, vampirical theory, and other concluding thoughts

The work that Podos defends is replete with problems that Andrew Gelman discusses on his website. I mention here just a few of the issues that particularly resonate with me:

" . . . the find-statistical-signficance-any-way-you-can-and-declare-victory paradigm."

" . . . what I've sometimes called the research incumbency rule: that, once an article is published in some approved venue, it should be taken as truth . . ."

" . . . the deadly combination of weak theory being supported almost entirely by statistically significant results which themselves are the product of uncontrolled researcher degrees of freedom."

" . . . huge, obvious multiple comparisons problems. . ."

"the real 'conclusion of the paper' doesn't depend on any of its details—all that matters is that there's *something*, somewhere, that has p less than .05, because that's enough to make publishable, promotable claims about 'the pervasiveness and persistence of . . .' whatever . . . they want to publish that day. When the authors protest that none of the errors really matter, it makes you realize that, in these projects, the data hardly matter at all."

" . . . the paradigm of the open-ended theory, of publication in top journals and promotion in the popular and business press, based on 'p less than .05' results obtained using abundant researcher degrees of freedom. It's the paradigm of the theory that . . . is 'more vampirical than empirical—unable to be killed by mere data'."

Abundant examples of "the garden of forking paths"

no acknowledgement of the simplest of alternative explanations for data, and instead a repeated effort to point out how consistent data are with a chosen explanation, which not coincidentally happens to be the preconceived idea that was in search of more evidence. " . . . the data hardly matter . . ."

Please cite this article in press as: Podos, J., Birdsong performance studies: reports of their death have been greatly exaggerated, Animal Behaviour (2016), http://dx.doi.org/10.1016/j.anbehav.2016.12.010

Above quotes from Andrew Gelman blog: http://andrewgelman.com/2016/09/21/what-has-happened-down-here-is-the-winds-have-changed/ ; and http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf

Dozens of colleagues, including many current colleagues at the University of Massachusetts Amherst, provided invaluable support, guidance and words of wisdom during the years that preceded this exchange. A special tip of the hat to Beth Jakob, Luke Remage-Healey, Chris Woodcock, Gordon Wyse, Melinda Novak, Duncan Irschick, Michelle Scott, Marc Naguib, Steve Nowicki, Bernie Lohr, Melissa Hughes, Sue Anne Zollinger, Henrik Brumm, Regina Macedo, Mario Cohn-Haft, Rick Prum, Lilian Manica, Sandy Vehrencamp, and, most of all, Cristina Cox Fernandes. At UMass I have been privileged to supervise the dissertation work of numerous first-rate young scientists, prominent among them Sarah Goodwin and Dana Moseley. For suffering through early drafts of this manuscript, I offer sincere thanks to Bill Searcy, Sandy Vehrencamp, Dan Mennill, Luke Remage-Healey, Sarah Goodwin, Dana Moseley, Beth Jakob, Norman Johnson, Dave Wilson and Melissa Hughes. Mike Beecher, David Logue and two anonymous referees provided sage advice and suggestions that sent me back to the drawing board; for their help I am grateful.

37) Acknowledgments—would someone please step forward?

I have no doubt that dozens of people offered moral support (three or four people were helping me); I think that, at one time or another, all of Podos' people were trying to undermine my message everywhere I turned (see my website). If these dozens of people all believed Podos was actually doing good science, then I despair.

Perhaps there's one way to find out. I suggest that Podos ask these dozens of people if they'd step forward and somehow declare, perhaps on Gelman's blog, that they fully support the science done by Podos, that the work done by Podos is of similar quality to their own. I would especially challenge the University of Massachusetts administrators to step forward and tell that this is how billions of dollars of federal grant money is spent at the University.

I might add that I am not totally heartless, a few opinions to the contrary, and I feel very bad for the students that Podos has put in the crosshairs of these discussions about science and ethics.

References

Akçay, Ç., & Beecher, M. D. (2015). Team of rivals in chipping sparrows? A comment on Goodwin & Podos. Biology Letters, 11, 20141043.

Arnold, S. J. (1983). Morphology, performance, and fitness. American Zoologist, 23(2), 347e361.

Ballentine, B. (2009). The ability to perform physically challenging songs predicts age and size in male swamp sparrows, Melospiza georgiana. Animal Behaviour, 77, 973e978. http://dx.doi.org/10.1016/j.anbehav.2008.12.027.

Ballentine, B., Hyman, J., & Nowicki, S. (2004). Vocal performance influences female response to male bird song: An experimental test. Behavioral Ecology, 15(1), 163e168. http://dx.doi.org/10.1093/beheco/arg090.

Barske, J., Fusani, L., Wikelski, M., Feng, N. Y., Santos, M., & Schlinger, B. A. (2014). Energetics of the acrobatic courtship in male golden-collared manakins

38) The Bottom Line, the Consequences of all this? The stakes are high, both personally and for science

On Andrew Gelman's blog, Podos writes the following (my emphasis):

> *Stepping back for a moment, and as a bit of a public service announcement, Kroodsma's stubborn adherence to contrarian viewpoints here comes as no surprise to veterans of our field. As Vehrencamp et al. (2017) observed, Kroodsma has previously applied the same modus operandi in attempts to discredit other major topics in the field of avian vocal communication. Of course there's nothing wrong with reasonable, fair criticism, as noted eloquently by Vehrencamp et al. in their opening paragraph. Yet* ***Kroodsma's critiques have moved well beyond reasonable and fair****. So I offer a message to future students stumbling on this thread;* ***take anything Kroodsma has published, critiques and science alike, with a heavy dose of skepticism****. But then again, you should also be highly critical of anything anyone writes, including what I am writing now on this non-peer-reviewed blog. Just take a look at the published exchange and relevant papers for yourselves, and take from it what you can — regardless of all the heat and noise.*

Podos puts the stakes pretty high:

What if Podos is right? Question everything Kroodsma has published, "critiques and science alike," says Podos

1) Then shame on Kroodsma for all the grief he has caused over the years. The motivation and substance of everything he has published during his career should be questioned (begin your questioning at http://donaldkroodsma.com/). (Podos, above: "*take* ***anything Kroodsma has published, critiques and science alike, with a heavy dose of skepticism, . . .***").

2) The science of birdsong and sexual selection is strong, as illustrated by the work of Podos and his colleagues, and is a premier illustration of how good science is done.

And if Kroodsma is right? Question everything Podos has published

1) Shame on Jeffrey Podos (and colleagues) for what he has published as science over the years. The motivation and substance of everything Podos has published during his career should be questioned (begin your questioning at his website at the University of Massachusetts, Amherst).

2) The literature I have critiqued, as illustrated by that of Podos and his colleagues, is a series of just-so stories with little to no truth in them, and is a premiere illustration of how science can go all wrong.

(Manacus vitellinus). Proceedings of the Royal Society B: Biological Sciences, 281(1776). http://dx.doi.org/10.1098/rspb.2013.2482.

Barske, J., Schlinger, B. A., & Fusani, L. (2015). The presence of a female influences courtship performance of male manakins. Auk, 132(3), 594e603. http://dx.doi.org/10.1642/auk-14-92.1.

Barske, J., Schlinger, B. A., Wikelski, M., & Fusani, L. (2011). Female choice for male motor skills. Proceedings of the Royal Society B: Biological Sciences, 278(1724), 3523e3528. http://dx.doi.org/10.1098/rspb.2011.0382.

Bateson, M., & Healy, S. D. (2005). Comparative evaluation and its implications for mate choice. Trends in Ecology & Evolution, 20(12), 659e664. http://dx.doi.org/10.1016/j.tree.2005.08.013.

Byers, B. E. (2007). Extrapair paternity in chestnut-sided warblers is correlated with consistent vocal performance. Behavioral Ecology, 18(1), 130e136.

Byers, J., Hebets, E., & Podos, J. (2010). Female mate choice based upon male motor performance. Animal Behaviour, 79, 771e778. http://dx.doi.org/10.1016/j.anbehav.2010.01.009.

Cardoso, G. C. (2013). Sexual signals as advertisers of resistance to mistakes. Ethology, 119(12), 1035e1043. http://dx.doi.org/10.1111/eth.12165.

Cardoso, G. C. (2017). Advancing the inference of performance in birdsong. Animal Behaviour, 125, e29ee32.

Cardoso, G. C., Atwell, J. W., Ketterson, E. D., & Price, T. D. (2007). Inferring performance in the songs of dark-eyed juncos (Junco hyemalis). Behavioral Ecology, 18(6), 1051e1057. http://dx.doi.org/10.1093/beheco/arm078.

Cardoso, G. C., Atwell, J. W., Ketterson, E. D., & Price, T. D. (2009). Song types, song performance, and the use of repertoires in dark-eyed juncos (Junco hyemalis). Behavioral Ecology, 20(4), 901e907. http://dx.doi.org/10.1093/beheco/arp079.

Caro, S. P., Sewall, K. B., Salvante, K. G., & Sockman, K. W. (2010). Female Lincoln's sparrows modulate their behavior in response to variation in male song quality. Behavioral Ecology, 21(3), 562e569. http://dx.doi.org/10.1093/beheco/arq022.

Doucet, S. M., & Montgomerie, R. (2003). Multiple sexual ornaments in satin bowerbirds: Ultraviolet plumage and bowers signal different aspects of male quality. Behavioral Ecology, 14, 503e509.

Draganoiu, T. I., Nagle, L., & Kreutzer, M. (2002). Directional female preference for an exaggerated male trait in canary (Serinus canaria) song. Proceedings of the Royal Society B: Biological Sciences, 269(1509), 2525e2531. http://dx.doi.org/10.1098/rspb.2002.2192.

DuBois, A. L., Nowicki, S., & Searcy, W. A. (2009). Swamp sparrows modulate vocal performance in an aggressive context. Biology Letters, 5(2), 163e165. http://dx.doi.org/10.1098/rsbl.2008.0626.

DuBois, A. L., Nowicki, S., & Searcy, W. A. (2011). Discrimination of vocal performance by male swamp sparrows. Behavioral Ecology and Sociobiology, 65(4), 717e726. http://dx.doi.org/10.1007/s00265-010-1073-2.

Forstmeier, W., Kempenaers, B., Meyer, A., & Leisler, B. (2002). A novel song parameter correlates with extra-pair paternity and reflects male longevity. Proceedings of the Royal Society B: Biological Sciences, 269(1499), 1479e1485. http://dx.doi.org/10.1098/rspb.2002.2039.

Garland, T. J., & Losos, J. B. (1994). Ecological morphology of locomotor performance in squamate reptiles. In P. C. Wainwright, & S. M. Reilly (Eds.), Ecological morphology: Integrative organismal biology (pp. 240e302). Chicago, IL: University of Chicago Press.

Geberzahn, N., & Aubin, T. (2014). Assessing vocal performance in complex birdsong: A novel approach. BMC Biology, 12, 9. http://dx.doi.org/10.1186/s12915-014-0058-4.

Gibson, J. S., & Uetz, G. W. (2008). Seismic communication and mate choice in wolf spiders: Components of male seismic signals and mating success. Animal Behaviour, 75, 1253e1262.

Goodwin, S. E., & Podos, J. (2014). Team of rivals: Alliance formation in territorial songbirds is predicted by vocal signal structure. Biology Letters, 10(2), 20131083. http://dx.doi.org/10.1098/rsbl.2013.1083.

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.

Hof, D., & Podos, J. (2013). Escalation of aggressive vocal signals: A sequential playback study. Proceedings of the Royal Society B: Biological Sciences, 280(1768), 8. http://dx.doi.org/10.1098/rspb.2013.1553.

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(1596), 1907e1912. http://dx.doi.org/10.1098/rspb.2006.3535.

Irschick, D. J. (2003). Measuring performance in nature: Implications for studies of fitness within populations. Integrative and Comparative Biology, 43(3), 396e407. http://dx.doi.org/10.1093/icb/43.3.396.

Irschick, D. J., Briffa, M., & Podos, J. (2015). Introduction. In D. J. Irschick, M. Briffa, & J. Podos (Eds.), Animal signaling and function: An integrative approach (pp. 1e9). Hoboken, NJ: John Wiley.

Irschick, D. J., Meyers, J. J., Husak, J. F., & Le Galliard, J.-F. (2008). How does selection operate on whole-organism functional performance capacities? A review and synthesis. Evolutionary Ecology Research, 10(2), 177e196.

Johnstone, R. A. (1997). The evolution of animal signals. In J. R. Krebs, & N. B. Davies (Eds.), Behavioural ecology (pp. 155e178). Oxford, U.K.: Blackwell.

de Kort, S. R., Eldermire, E. R. B., Cramer, E. R. A., & Vehrencamp, S. L. (2009). The deterrent effect of bird song in territory defense. Behavioral Ecology, 20(1), 200e206. http://dx.doi.org/10.1093/beheco/arn135.

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

Lahti, D. C., Moseley, D. L., & Podos, J. (2011). A tradeoff between performance and accuracy in bird song learning. Ethology, 117(9), 802e811. http://dx.doi.org/10.1111/j.1439-0310.2011.01930.x.

Lambrechts, M. M. (1996). Organization of birdsong and constraints on performance. In D. Kroodsma, & E. Miller (Eds.), Ecology and evolution of acoustic communication in birds (pp. 305e320). Ithaca, NY: Comstock.

Liu, W.-C., & Kroodsma, D. E. (1999). Song development by chipping sparrows and field sparrows. Animal Behaviour, 57, 1275e1286. http://dx.doi.org/10.1006/anbe.1999.1081.

Liu, W.-C., & Kroodsma, D. E. (2006). Song learning by chipping sparrows: When, where, and from whom. Condor, 108(3), 509e517. http://dx.doi.org/10.1650/0010-5422(2006)108[509:slbcsw]2.0.co;2.

Liu, W.-C., & Nottebohm, F. (2007). A learning program that ensures prompt and versatile vocal imitation. Proceedings of the National Academy of Sciences of the United States of America, 104(51), 20398e20403. http://dx.doi.org/10.1073/pnas.0710067104.

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(1), 34e41. http://dx.doi.org/10.1086/587849.

Lyons, S. M., Beaulieu, M., & Sockman, K. W. (2014). Contrast influences female attraction to performance-based sexual signals in a songbird. Biology Letters, 10(10). http://dx.doi.org/10.1098/rsbl.2014.0588.

Marler, P. (1997). Three models of song learning: Evidence from behavior. Journal of Neurobiology, 33(5), 501e516.

Marler, P., & Hamilton, W. J. (1966). Mechanisms of animal behavior. New York, NY: John Wiley.

Moseley, D. L., Lahti, D. C., & Podos, J. (2013). Responses to song playback vary with the vocal performance of both signal senders and receivers. Proceedings of the Royal Society B: Biological Sciences, 280(1768), 20131401. http://dx.doi.org/10.1098/rspb.2013.1401.

Naguib, M., Schmidt, R., Sprau, P., Roth, T., Floercke, C., & Amrhein, V. (2008). The ecology of vocal signaling: Male spacing and communication distance of different song traits in nightingales. Behavioral Ecology, 19(5), 1034e1040. http://dx.doi.org/10.1093/beheco/arn065.

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. http://dx.doi.org/10.1016/j.anbehav.2011.04.018.

Patricelli, G. L., & Krakauer, A. H. (2009). Tactical allocation of effort among multiple signals in sage grouse: An experiment with a robotic female. Behavioral Ecology, 21, 97e106. http://dx.doi.org/10.1093/beheco/arp155.

Podos, J. (1996). Motor constraints on vocal development in a songbird. Animal Behaviour, 51, 1061e1070.

Podos, J., Lahti, D. C., & Moseley, D. L. (2009). Vocal performance and sensorimotor learning in songbirds. Advances in the Study of Behavior, 40, 159e195.

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.

Podos, J., & Patek, S. N. (2015). Acoustic signal evolution: Biomechanics, size, and performance. In D. J. Irschick, M. Briffa, & J. Podos (Eds.), Animal signaling and function: An integrative approach (pp. 175e203). Hoboken, NJ: John Wiley.

Podos, J., Peters, S., & Nowicki, S. (2004). Calibration of song learning targets during vocal ontogeny in swamp sparrows, Melospiza georgiana. Animal Behaviour, 68, 929e940.

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(4), 425e437. http://dx.doi.org/10.1086/667575.

Sakata, J. T., & Vehrencamp, S. L. (2012). Integrating perspectives on vocal performance and consistency. Journal of Experimental Biology, 215(2), 201e209. http://dx.doi.org/10.1242/jeb.056911.

Searcy, W. A., & Beecher, M. D. (2009). Song as an aggressive signal in songbirds. Animal Behaviour, 78, 1281e1292. http://dx.doi.org/10.1016/j.anbehav.2009.08.011.

Searcy, W. A., & Nowicki, S. (2005). The evolution of animal communication: Reliability and deception in signaling systems. Princeton, NJ: Princeton University Press. Suthers, R. A., Vallet, E., & Kreutzer, M. (2012). Bilateral coordination and the motor basis of female preference for sexual signals in canary song. Journal of Experimental Biology, 215(17), 2950e2959. http://dx.doi.org/10.1242/jeb.071944.

Vehrencamp, S. L., de Kort, S. R., & Illes, A. E. (2017). Response to Kroodsma's critique of banded wren song performance research. Animal Behaviour, 125, e25ee28.

Wainwright, P. C. (1994). Functional morphology as a tool in ecological research. In P. C. Wainwright, & S. M. Reilly (Eds.), Ecological morphology: Integrative organismal biology (pp. 42e59). Chicago, IL: University of Chicago Press.

Welch, A. M., Semlitsch, R. D., & Gerhardt, H. C. (1998). Call duration as an indicator of genetic quality in male gray tree frogs. Science, 280, 1928e1930.

Wiley, R. H. (2003). Is there an ideal behavioural experiment? Animal Behaviour, 66, 585e588. http://dx.doi.org/10.1006/anbe.2003.2231.

Wilson, D. R., Bitton, P. P., Podos, J., & Mennill, D. J. (2014). Uneven sampling and the analysis of vocal performance constraints. American Naturalist, 183(2), 214e228. http://dx.doi.org/10.1086/674379.

Zollinger, S. A., Podos, J., Nemeth, E., Goller, F., & Brumm, H. (2012). On the relationship between, and measurement of, amplitude and frequency in birdsong. Animal Behaviour, 84. http://dx.doi.org/10.1016/j.anbehav.2012.04.026.

Zollinger, S. A., & Suthers, R. A. (2004). Motor mechanisms of a vocal mimic: Implications for birdsong production. Proceedings of the Royal Society B: Biological Sciences, 271(1538), 483e491. http://dx.doi.org/10.1098/rspb.2003.2598.

Please cite this article in press as: Podos, J., Birdsong performance studies: reports of their death have been greatly exaggerated, Animal Behaviour (2016), http://dx.doi.org/10.1016/j.anbehav.2016.12.010