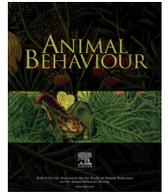




Contents lists available at ScienceDirect

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.

ARTICLE INFO

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, e1–e16) 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.

© 2016 Published by Elsevier Ltd on behalf of The Association for the Study of Animal Behaviour.

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, Kempnaers, 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.

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.*

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

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?

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 5–7 '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).

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.

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 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.

To be clear, we still have much to learn about whether and how 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. 181–182) 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.

Swamp Sparrows

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 6–8). Thus, as with chipping sparrows, Kroodsma questions whether young male swamp sparrows could ever tailor songs to their own individual proficiency, given the apparent premium on song learning accuracy.

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.

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. e8–e12). 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 individual birds' 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). 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 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.

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 Kroodasma's declaration that 'an important condition for honesty and reliability is that males consistently use songs within a relatively narrow range of performance abilities' (Kroodasma, 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 Kroodasma. 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, Kroodasma's blanket dismissal of this area of research ('the hypothesis becomes biologically implausible, if not impossible'; Kroodasma, 2017, page e1) is surely premature.

VOCAL PERFORMANCE AND SONG FUNCTION

Kroodasma'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 Kroodasma instead offers are alternative hypotheses that might also explain the outcome. I regard two of these as credible and worth further attention. First, Kroodasma 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.

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

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).

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. 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). Finally, it is worth noting that birds' responses to the most highly manipulated songs, even at the lowest trill rates, tended 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, 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. 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). 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).

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).

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. 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). 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. 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

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.

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.

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.

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.

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. Where in [DuBois et al. \(2009, 2011\)](#) Kroodsma sees conceptual flaws, I see the application 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. Kroodsma also notes that the magnitude 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.

[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.

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

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 provisional basis. Consistency of data with a hypothesis does not imply the hypothesis ‘must therefore be true’ (Kroodsma, 2017, page e7).

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.

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.

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 natural settings are able to develop songs with faster trill 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.

Acknowledgments

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 Ramage-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 Ramage-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.

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), 347–361.
- Ballentine, B. (2009). The ability to perform physically challenging songs predicts age and size in male swamp sparrows, *Melospiza georgiana*. *Animal Behaviour*, *77*, 973–978. <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), 163–168. <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

- (*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), 594–603. <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), 3523–3528. <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), 659–664. <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), 130–136.
- Byers, J., Hebets, E., & Podos, J. (2010). Female mate choice based upon male motor performance. *Animal Behaviour*, 79, 771–778. <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), 1035–1043. <http://dx.doi.org/10.1111/eth.12165>.
- Cardoso, G. C. (2017). Advancing the inference of performance in birdsong. *Animal Behaviour*, 125, e29–e32.
- 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), 1051–1057. <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), 901–907. <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), 562–569. <http://dx.doi.org/10.1093/beheco/arg022>.
- 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, 503–509.
- 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), 2525–2531. <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), 163–165. <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), 717–726. <http://dx.doi.org/10.1007/s00265-010-1073-2>.
- Forstmeier, W., Kempnaers, B., Meyer, A., & Leisler, R. (2002). A novel song parameter correlates with extra-pair paternity and reflects male longevity. *Proceedings of the Royal Society B: Biological Sciences*, 269(1499), 1479–1485. <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. 240–302). Chicago, IL: University of Chicago Press.
- Geberzahn, N., & Aubin, T. (2014). Assessing vocal performance in complex bird-song: 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, 1253–1262.
- 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), 1907–1912. <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), 396–407. <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. 1–9). 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), 177–196.
- Johnstone, R. A. (1997). The evolution of animal signals. In J. R. Krebs, & N. B. Davies (Eds.), *Behavioural ecology* (pp. 155–178). 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), 200–206. <http://dx.doi.org/10.1093/beheco/arn135>.
- Kroodsma, D. (2017). Birdsong performance studies: A contrary view. *Animal Behaviour*, 125, e1–e16.
- Lahti, D. C., Moseley, D. L., & Podos, J. (2011). A tradeoff between performance and accuracy in bird song learning. *Ethology*, 117(9), 802–811. <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. 305–320). Ithaca, NY: Comstock.
- Liu, W.-C., & Kroodsma, D. E. (1999). Song development by chipping sparrows and field sparrows. *Animal Behaviour*, 57, 1275–1286. <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), 509–517. [http://dx.doi.org/10.1650/0010-5422\(2006\)108\[509:sblscsw\]2.0.co;2](http://dx.doi.org/10.1650/0010-5422(2006)108[509:sblscsw]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), 20398–20403. <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), 34–41. <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), 501–516.
- 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), 1034–1040. <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, 177–183. <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, 97–106. <http://dx.doi.org/10.1093/beheco/arp155>.
- Podos, J. (1996). Motor constraints on vocal development in a songbird. *Animal Behaviour*, 51, 1061–1070.
- Podos, J., Lahti, D. C., & Moseley, D. L. (2009). Vocal performance and sensorimotor learning in songbirds. *Advances in the Study of Behavior*, 40, 159–195.
- 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, 203–212.
- 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. 175–203). 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, 929–940.
- 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), 425–437. <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), 201–209. <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, 1281–1292. <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), 2950–2959. <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, e25–e28.
- 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. 42–59). 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, 1928–1930.
- Wiley, R. H. (2003). Is there an ideal behavioural experiment? *Animal Behaviour*, 66, 585–588. <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), 214–228. <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), 483–491. <http://dx.doi.org/10.1098/rspb.2003.2598>.