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

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

1) The calibration hypothesis was supported by two lines of evidence . . .

2) . . . in directions consistent with expectations [of the calibration hypothesis]

3) Consistent with the calibration hypothesis, we found . . .

4) The direction of this transition is consistent with the calibration hypothesis

5) The observation is consistent with our expectation . . .

6) Consistent with this result, we also found that differences in trill rates . . .

7) The main support for the calibration hypotheses comes from analysis of . . .

8) . . . songs more consistent with expected levels of proficiency

9) Another line of evidence that provides support for the calibration hypothesis . . .

10) This bias is consistent with the calibration hypothesis . . .

11) A third line of evidence . . . offers additional . . . support for the calibration hypothesis

12) These results were consistent with the calibration hypothesis

I provide excerpts from all examples, because all together they illustrate how selective authors can be in reporting evidence to support the performance hypothesis. The impression conveyed is that, given how much is consistent with the calibration hypothesis, it must therefore be true. But the words "consistent with" are also red flags for readers to ask what other hypotheses the data might be consistent with.

The figures I provide on the biology of swamp sparrow song soundly reject the authors' conclusions about calibration. When learning a song, a swamp sparrow conforms to the particular song type of the local dialect; he does not adjust features of what he learns in any way consistent with an attempt to calibrate a song to his own abilities. As a result, a male may have the worst "performance" on one song type, the best on another.

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

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

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

> Science is the search for truth, regardless of how good the story is; "marketing and advertising" are the search for a good story, regardless of the truth (October 2004)

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

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

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

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

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

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

> Our main finding is that birds elevated the trill rates of low-performance models, 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 (*Podos et al. 2004, 2009*) . . . Prior work in swamp sparrows showed calibration . . . (*Podos et al. 2004*).

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

I struggle with these statements. What is certainly true is that the young swamp sparrows removed silent intervals from odd, slowed tutor songs to produce more normal, wild-type songs. That result, however, based on abnormal, experimental songs, does not warrant any conclusion about a young swamp sparrow either in nature or in the laboratory taking a natural tutor song that it hears, foregoing "imitative accuracy," and adjusting that song in trill rate or frequency bandwidth to match his own capabilities, all so that he can honestly broadcast his individual quality. There are no data in this paper or elsewhere demonstrating that a young swamp sparrow adjusts a normal or abnormal song to match his own individual proficiency, only data showing

how young birds strive to produce normal, species-typical songs. There are no data that demonstrate any honesty in the signaling of trill rate or frequency bandwidth.

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

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

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

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

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

The two confounding variables are a serious problem. First, I think it not unreasonable to expect that a "normal" song with three times as much stimulus as the lowest performance song might elicit a stronger response, based on *stimulus quantity* alone. But never once do the authors offer any control for or thought about such a simple, obvious alternative explanation for their results. Without *somehow* controlling for this confounding factor, one cannot attribute response strength to trill rate alone.

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

abnormal as to not be biologically meaningful; nevertheless, Moseley et al. use songs well outside that range.

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

> "territorial male swamp sparrows responded significantly less strongly to low-performance [abnormal] than to control-performance [normal] playback stimuli, consistent with our hypothesis [any hypothesis] that receivers should attribute limited threat to low-performance [abnormal] songs . . .

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

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

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

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

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

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

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

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

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