



CHICAGO JOURNALS



The University of Chicago

How Reliable Is Song Learning Accuracy as a Signal of Male Early Condition?

Author(s): R. F. Lachlan and S. Nowicki

Reviewed work(s):

Source: *The American Naturalist*, Vol. 180, No. 6 (December 2012), pp. 751-761

Published by: [The University of Chicago Press](#) for [The American Society of Naturalists](#)

Stable URL: <http://www.jstor.org/stable/10.1086/668010>

Accessed: 08/01/2013 16:28

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



The University of Chicago Press, The American Society of Naturalists, The University of Chicago are collaborating with JSTOR to digitize, preserve and extend access to *The American Naturalist*.

<http://www.jstor.org>

How Reliable Is Song Learning Accuracy as a Signal of Male Early Condition?

R. F. Lachlan* and S. Nowicki

Department of Biology, Duke University, Box 90338, Durham, North Carolina 27708

Submitted February 1, 2012; Accepted July 5, 2012; Electronically published October 26, 2012

Online enhancements: appendices.

ABSTRACT: That many species of songbirds learn their songs imitatively is well established, but it is less clear why they do so. A component of the developmental-stress hypothesis posits that young males in good condition learn songs more accurately than males in poor condition and that females use learning accuracy as an honest signal of male developmental history. An unresolved problem is how females reliably assess learning accuracy when they are not certain of the identity of the male's tutor and thus the specific model from which a song was copied. We therefore investigated whether song learning accuracy assessment (SLAA) can be reliable, using evolutionary simulation models of song learning. We found that SLAA is indeed less reliable than assessment in which male signals are compared to an unlearned standard, as a result of three types of errors in matching songs to their models. In the simplest models, SLAA was particularly unreliable, but when the model is made more realistic by including features such as geographically constrained learning, repertoire complexity, and, in particular, song categorization, the reliability of SLAA increased. Our results demonstrate a range of conditions under which the assessment of song learning accuracy might be reasonably reliable and therefore likely to evolve.

Keywords: developmental stress hypothesis, signaling system, cultural evolution, song learning, assessment, simulation models.

Introduction

Many songbirds learn their songs by imitating conspecifics, as demonstrated by hand-rearing experiments in the laboratory (e.g., Thorpe 1958; Marler and Peters 1981; Tchernichovski et al. 2001) and inferred from the high proportion of individuals singing shared song types in a broad range of species in the field (see, e.g., Podos and Warren 2007 and Catchpole and Slater 2008 for reviews; see Lynch 1996 and Lachlan and Slater 2003 for quantitative analyses). Of the major hypotheses that explain why birds might learn songs from others, the most recent is a component of the developmental-stress hypothesis (Nowicki

et al. 1998, 2002a; Searcy and Nowicki 2005). Developmental stress experienced early in life has long-lasting effects on brain development in songbirds, including on those brain areas involved with song learning and production (e.g., Nowicki et al. 2002a; Buchanan et al. 2004; MacDonald et al. 2006). The developmental-stress hypothesis proposes that song may serve as an intersexual assessment signal because young males that experience less stress and are thus able to invest more resources in brain development are also expected to be correspondingly better singers, potentially making song a reliable indicator of male quality.

One aspect of song likely to be influenced by brain development is how accurately a male imitates the songs he attempts to copy. If males in poor condition learn song less accurately than males in good condition, then females may assess male condition through the proxy of song learning accuracy. We call this component of the developmental-stress hypothesis "song learning accuracy assessment" (SLAA). SLAA has received two types of empirical support. First, males do indeed learn songs less precisely if raised under developmental stress (swamp sparrows *Melospiza georgiana*: Nowicki et al. 2002a; zebra finches *Taeniopygia guttata*: Holveck et al. 2008; Brumm et al. 2009; but see Gil et al. 2006; Zann and Cash 2008), demonstrating the potential for learning accuracy to provide information about males to receivers. Second, and more critically, females show diminished sexual response to inaccurately learned songs (song sparrows *Melospiza melodia*: Nowicki et al. 2002b; swamp sparrows: R. F. Lachlan, R. C. Anderson, and S. Nowicki, unpublished data), consistent with the interpretation that they use this signal in the context of mate choice to assess a male's early condition.

Perhaps the biggest challenge for the SLAA hypothesis is to explain how females are able to assess male song learning accuracy, which requires them to have some point of comparison. Because young males may learn their songs from different tutors, however, and because songs often

* Corresponding author; e-mail: rfl5@duke.edu.

Am. Nat. 2012. Vol. 180, pp. 751–761. © 2012 by The University of Chicago. 0003-0147/2012/18006-53622\$15.00. All rights reserved.

DOI: 10.1086/668010

vary greatly among adult males in a population, it is not clear how a female can tell which song or songs a particular male attempted to learn. Given this problem, we can identify four alternative mechanisms by which a female might assess how well a male has learned his songs.

Two mechanisms circumvent the problem entirely. First, a female might simply know the identity of a male's tutor and have memorized the tutor's songs herself, thus giving her direct knowledge of the specific songs that male attempted to copy. In most species, however, males do not learn from any predictably identifiable individual (reviewed in Catchpole and Slater 2008). In a few species, males do appear to learn from their fathers (e.g., Grant and Grant 1996), but even in this special case it is not clear how a female could know the identity of a particular male's father (unless they were nestmates, in which case SLAA would conflict with inbreeding avoidance). We therefore deem it unlikely that this mechanism would apply in the real world. A second mechanism was modeled by Ritchie et al. (2008), who posited that some features of song can be accurately imitated only by high-quality males and that these features (e.g., greater length) are assessed by females in lieu of learning accuracy itself. While this is a plausible assessment mechanism, we do not think that it ultimately provides an explanation for accurate learning by males: if a high-quality male attempted to learn from a low-quality tutor, there would inevitably be a conflict between learning accuracy and producing an attractive song (Lahti et al. 2011), and selection would then favor males that maximized the preferred feature at the expense of accurate learning.

Two other, more promising mechanisms for assessing song learning accuracy involve different ways in which a female might infer the model for a male's song, either by comparing his song to specific song exemplars she has memorized or by comparing it to generalized song-type categories she has learned. In this article, we use evolutionary models to examine the plausibility of these two mechanisms. In so doing, we investigate the conditions under which SLAA can be reliable enough to evolve and be used in mate choice as an indicator of early developmental stress.

In the first mechanism, a female attempts to compare a male's song to the best match among a collection of specific songs she herself has memorized previously from the population. Such "inference from exemplars" seems straightforward at first, but three different types of mistakes can occur. (1) A female might never have encountered the model song from which the male she is assessing learned; we call this "novel-song error" (fig. 1A). This may happen if the male learned his song outside the female's geographical range, for example, or if the male is older than the female and his tutor had died before she had the

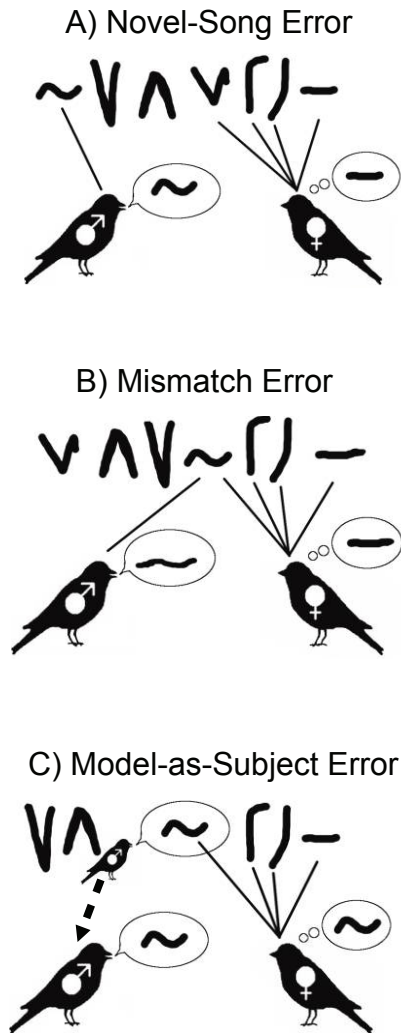


Figure 1: Three sources of error particular to song learning accuracy assessment. Each figure shows a range of tutor signals along the top and how a male and a female might learn from those models. *A*, Novel-song error. The female does not learn the model that the male bases his song on and compares his song with a different model instead. *B*, Mismatch error. The male learns his song with enough inaccuracy that the female matches his song to the wrong model, even though she has learned the correct model. *C*, Model-as-subject error. The male is actually a member of the tutor cohort, and the female learns his song before later assessing him as a potential mate. This artificially inflates her judgment of his song learning accuracy.

chance to be exposed to that tutor's songs. (2) A female might have learned the correct model song but might match the male song to a different, false model; we call this "mismatch error" (fig. 1B). This may occur if the male learned his song inaccurately enough that it actually resembles the false model more closely than the true model. (3) If a female reencounters a male that she had learned from previously (e.g., in the previous year), she might

attempt to match his song to her memory of his song; we call this “model-as-subject error” (fig. 1C). Unlike novel-song errors, the consequence of a mismatch error or a model-as-subject error is that the female assesses the male to have learned his song more accurately than he actually did. In our simplest simulations, we find that the potential for these kinds of mistakes makes SLAA notably unreliable, as compared to an unlearned assessment signal. Reliability increases dramatically, however, as we incorporate realistic features into the model, such as geographically constrained learning and repertoire complexity.

In the second mechanism that we examine, females assess male songs not by comparing them to specific exemplars they have memorized but rather on the basis of their adherence to learned population-wide categories of song types. That is, we assume that a song is assessed by the degree to which it is prototypical of the category to which it belongs. Studies of category formation in human perception, including studies of human speech, support the idea that prototypes within categories may be ubiquitous. For example, human infants form prototypes of speech sounds during language learning (Grieser and Kuhl 1989). There is also evidence that animals form prototypes when they are trained to perceive human speech (Kluender et al. 1998). Category formation may allow birds to avoid the errors depicted in figure 1 because males and females may form similar population-specific prototypes for song-type categories even if the song exemplars they are exposed to come from different sets of tutors in that population (fig. 2). A human analogy to this is that two people might both have similar concepts of what constitutes a typical

New York English accent even if they have learned these concepts from hearing different sets of New Yorkers. When we add to our model an algorithm by which birds develop song categories generalized from the aggregate of song exemplars they are exposed to and females base their assessment against prototypical versions of a song-type category, we find that SLAA becomes highly reliable as an assessment strategy and thus likely to evolve.

Model 1: Inference of Song Learning Accuracy by Matching to Song Exemplars

The potential errors in matching a male’s song to exemplars in a female’s memory (fig. 1) can be overwhelming. In a simple, nonspatial mathematical model of the hypothesis, one can demonstrate that SLAA is unreliable to the point of providing nearly no information to the female at all (app. A, available online). In this section, we examine ways in which a more realistic model of the process of song learning might improve reliability. For example, novel-song errors (fig. 1A) should be reduced if males occupy a territory only a short distance from where they learned their songs. There is evidence of such a geographical constraint to learning in many of the species of songbird studied to date: song types are shared most commonly with neighbors or nearby individuals (see Podos and Warren 2007 for a recent review).

In the real world, birds often sing more than one song type, and the songs themselves are also normally complex, containing many discrete elements. While, in some species, there may be little organization to the sequencing of songs

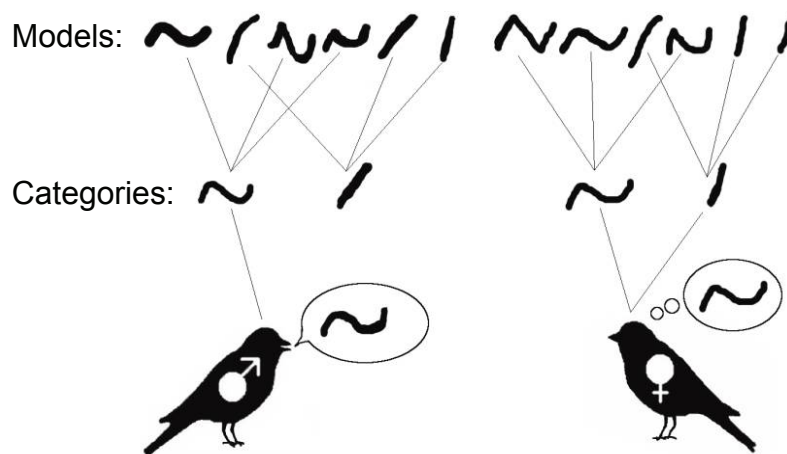


Figure 2: How categorical perception might influence song learning and song preferences. Males and females first form categories from the signals they hear, and then females base their assessment on how closely a male’s songs match a prototypical version of the song type category it best fits. Thus, even if males and females hear a different set of tutor songs, male song targets might closely match female preferences. Here, the song preference that the female has matched the male song with is very similar to the category average that the male was trying to learn.

within a repertoire (e.g., zebra finches: Lachlan et al. 2010), in others there may also be “syntax” constraints on where certain types of elements or syllables can be placed within a song (e.g., white-crowned sparrows *Zonotrichia leucophrys*: Soha and Marler 2001). In some species, elements or syllables are learned from multiple songs and recombined into the songs in a male’s repertoire (e.g., eastern song sparrow populations: Hughes et al. 1998), while in others, entire songs are learned intact from their models (e.g., chaffinches *Fringilla coelebs*: Thorpe 1958; Lachlan and Slater 2003; western song sparrow populations: Beecher et al. 1994b), including the sequence of units. These different types of repertoire organization all influence the information received by a female assessing song learning accuracy, and they might also be expected to influence the reliability of assessment.

Methods

Our models use individual-based simulations in which each individual in the population occupies a position on a two-dimensional grid. During each time period (“year”) of the simulation, females choose mates and produce offspring. Male offspring then occupy territories left open when territory holders die. Males and females disperse up to $d_{i,m}$ and $d_{i,f}$ territories, respectively, to find a mate or an empty territory. The number of offspring produced by a pair is related to the condition of the male early in his life, q_m (a noninherited trait set randomly at hatching), so it is in a female’s interest to choose a male with a high q_m . The model is an example of a direct-benefits model of sexual selection (Kirkpatrick and Ryan 1991), a choice we made to generate a simple and stable selection pressure for mate assessment by females.

In most of our models, males and females learn songs in their first year of life from males within $d_{s,m}$ and $d_{s,f}$ territories of their natal territory, respectively. In many species in the real world, males learn after dispersing from their natal territory (reviewed in Podos and Warren 2007). In these cases, males first disperse to some area of suitable habitat, then learn their songs, and then search for an available territory nearby. We can modify our model to account for this behavior by adding an additional step of an initial dispersal phase, occurring before song learning, in which males disperse up to a distance of d_{i,m_2} territories from their natal territory. After this, the model proceeds as before, with males learning songs from within $d_{s,m}$ territories of their new location and searching over $d_{i,m}$ territories for an open territory. When we modified our model to include this dispersal step, however, we found that it had no effect on the outcome of the simulations (using various values of d_{i,m_2}), so we do not report these results separately.

Songs are characterized as real-valued numbers between 0 and 1, arranged on a ring. That is, songs of values 1 and 0 are identical, and the similarity between two songs can be taken as the difference between their song values around the ring (this feature of the model avoided edge effects). We vary details of song learning in our different simulations (below), but in all cases, male song learning accuracy depends on male condition, q_m : songs are learned with an error randomly drawn from a normal distribution with variance $(1 - q_m)p_1$. The default value of p_1 in the simulations is 0.02.

In our simplest models, individuals have a repertoire, R , of just one element. When $R > 1$, we modeled three different types of complex repertoires. For males with “structureless” repertoires, repertoire elements are selected at random from the complete list of the repertoire elements of all the potential tutors. For males with “syntactic” repertoires, each element in a male’s repertoire is selected from the list of elements that occupied the same position in the tutors’ repertoires. Thus, in this model, the different elements in a male’s repertoire would, in most cases, be learned from different tutors but would all occupy the same position in the male’s repertoire as they did in the tutors’ repertoires. For males with “whole-repertoire learning,” one tutor is selected at random from the list of tutors, and his entire repertoire is learned, maintaining the position of elements within the repertoire.

Males also possess a nonlearned signal, which may stand, for example, for a visual signal such as plumage. This nonlearned signal also depends on males’ early condition: its value is $q_m + E(0, p_2)$. Here, $E(0, p_2)$ is a pseudorandom Gaussian error term that adds noise to the signal: the higher the variance term p_2 , the lower the correlation between the signal and the male’s condition.

Females could choose males on the basis of either of the two male signals. Which of the signals a female assesses depends on the inheritance of a single, autosomal gene with two alleles: a and A . The a females rely on the nonlearned signal, while the A females use SLAA to assess males. We examined the evolution of SLAA by investigating the probability of a population of a individuals being invaded by A individuals, when a small number (5% of the population) of A individuals were introduced to the population, and, conversely, the probability of a population of A individuals being invaded by a individuals.

Females using the nonlearned signal prefer the male with the highest signal value, while females using SLAA prefer the male singing a song most similar to a model that they had learned. When repertoire size, $R > 1$, female choice varied according to the type of repertoire. For structureless repertoires, females simply assessed each element in a male’s repertoire independently, without respect to relative position of the element. For syntactic-learning rep-

ertoires, females assessed each element in a male's repertoire by comparing it only with elements that they had learned that also occurred in the same position within the tutor's repertoire. For both the structureless-repertoire and the syntactic-learning models, the overall preference for a male was calculated as the average preference for each of the elements in his repertoire. Finally, for whole-song learning repertoires, females compared each element in a male's repertoire with the element in the same position of a tutor's repertoire. An overall score for each tutor was then calculated by averaging these element scores.

For each of our models, we estimated the reliability of the nonlearned signal that resulted in equal evolutionary success for the two signaling assessment strategies. To do this, we varied the parameter p_2 until the probability of a invading a population of A was equal to the probability of A invading a population of a . For this equilibrium value, \hat{p}_2 , we then estimated the correlation coefficient \hat{r}_v between male condition and signal value. This coefficient then served as our main measure of the evolvability and reliability of SLAA. A high value of \hat{r}_v meant that SLAA could outcompete a nonlearned signal even when there was a strong correlation between male condition and the nonlearned-signal value. We provide details of the implementation of the simulation in appendix B, available online.

Results

We first examined the reliability of SLAA with minimal spatial structure. We simplified our model by allowing females to learn only one song, setting R to 1, and setting the mortality rate, M , to 1, which effectively meant that males and females could breed during one year only, preventing overlap of generations. Then, when we set all dispersal distances to maximum values, we found that $\hat{r}_v = 0.006$. In other words, to make the SLAA strategy equally successful as the null, non-song learning strategy, we had to reduce the correlation between a male's condition and his nonlearned signal nearly to 0. This result confirms the result of our simple nonspatial mathematical model (app. A) and suggests that with moderate or larger dispersal distances, SLAA based on song exemplars is an inherently unreliable mechanism for assessing male condition.

We introduced spatial structure to the simulation by reducing dispersal distances. The effect of this was to increase \hat{r}_v , but as long as females learned only one song, \hat{r}_v remained low. The maximum value we found was 0.07, which occurred when males learned only from fathers and dispersed only to neighboring territories.

If females could learn all songs within $d_{s,f}$ territories, rather than just one, however, \hat{r}_v increased dramatically (fig. 3A). As expected, \hat{r}_v increased as the distance over

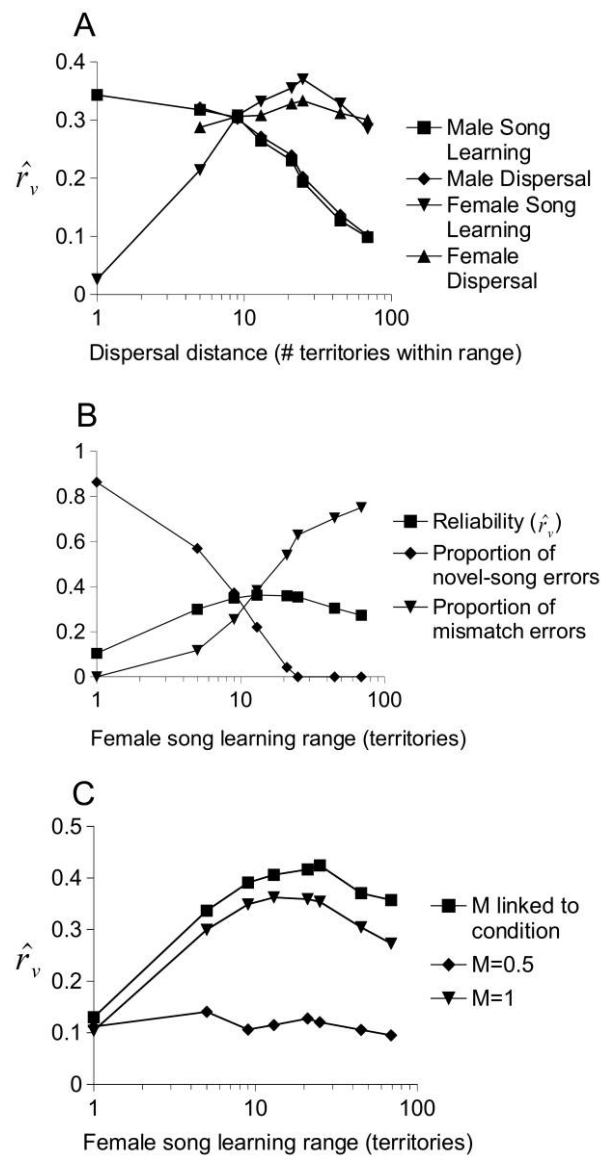


Figure 3: Song learning ranges and reliability. The plots show how varying the range over which males or females learn songs or disperse influences reliability. A, With $M = 1$ (i.e., no overlapping generations). B, How two different types of error combine to influence reliability (\hat{r}_v). Note that for B, the Y-axis shows both proportions and reliability scores on the same scale. C, With different patterns of mortality rate. With $M = 0.5$ or with M linked to condition, there are overlapping generations and the possibility of model-as-subject errors. In these simulations, females learned all songs within range, and unless otherwise stated, dispersal distances were 9 territories.

which males learned songs and dispersed decreased. This is an intuitive result because reducing male dispersal also reduces novel-song error (fig. 1A), by increasing the probability that females would be familiar with a male's tutor. A different pattern emerged for female song learning dis-

tance and female dispersal (fig. 3A): \hat{r}_v increased with $d_{s,f}$ until $d_{s,f}$ was large enough to encompass all the potential tutors of all potential mates that females might encounter. Beyond this point, reliability declined again. Overall, the highest level of \hat{r}_v that we measured with $M = 1$ was 0.439, which occurred when males learned songs only from one of the neighboring territories and when females learned all songs within a range of 2 territories but searched for mates only within 1 territory distance. As the number of males a female learns from increases, novel-song errors may decrease, but the probability of mismatch error increases (fig. 3B). The trade-off between these two types of error explains why reliability reached a maximum at an intermediate value of $d_{s,f}$ (fig. 3B).

We next considered the effect of allowing overlapping generations, that is, $M < 1$. This added a third source of error to our simulations: model-as-subject error (fig. 3C), where females mistakenly assess an older male as having accurately learned his own song. However, the extent to which this counts as an error depends on our assumptions: a male's age might indicate his condition and consequently his suitability as a mate. Females might, therefore, use familiarity with a song as a way of distinguishing older from first-year males. Under this interpretation, model-as-subject error is not so much an error as an alternative way of assessing a song. Empirical evidence suggests that females prefer older males in some, but not all, species (Brooks and Kemp 2001), so the frequency with which this situation really is an "error" remains unclear.

Because of the uncertainty about model-as-subject error, we investigated overlapping generations in two different ways. In the first, M was set to 0.5 per year for all individuals. In these simulations, there was no benefit for females to mate with older males. In the second, M was set to 0.5 for females, but for males, it was set to $M_m = 1 - (q_m/2)$; that is, males in better condition had lower mortality rates, and by choosing older males, females were therefore more likely to also choose a male in better condition.

As expected, these two conditions—linking and not linking male mortality to condition—had very different outcomes on the overall reliability of SLAA (fig. 3C). With condition-linked mortality ($M_m = 1 - (q_m/2)$), \hat{r}_v increased compared to that in our initial model, while otherwise ($M = 0.5$), \hat{r}_v decreased considerably. Model-as-subject errors therefore appeared to have a large effect on the outcome of our simulations. This makes sense in the model, because a model-as-subject error always vaults a previous tutor to the head of the list of potential mates, because his assessed score, t_m , is always at the maximum. Thus, model-as-subject errors always influenced mate choice decisions if they occurred. In reality, females might not be able to make such precise assessments, and in ad-

dition, males' songs might change slightly from year to year, so our models may overstate the importance of this type of error.

Increasing repertoire complexity also increased the reliability of SLAA, under all three conditions we examined (fig. 4). This is in line with the "backup" theory of signal evolution (Johnstone 1996): each additional repertoire component adds to the overall reliability of assessment. There were, however, clear differences between the three models: structureless repertoires led to the lowest reliability scores of the three, while whole-song learning led to the highest scores (fig. 4A). The root of this difference lies in the frequency of mismatch errors (fig. 4B). In an unstructured repertoire, the potential for mismatch error for each element in the repertoire increases with repertoire size because there are more potential tutor songs in a female's memory to be confused. In contrast, with a syntactically structured song learning strategy, each position in the repertoire is independent (somewhat as genetic loci on dif-

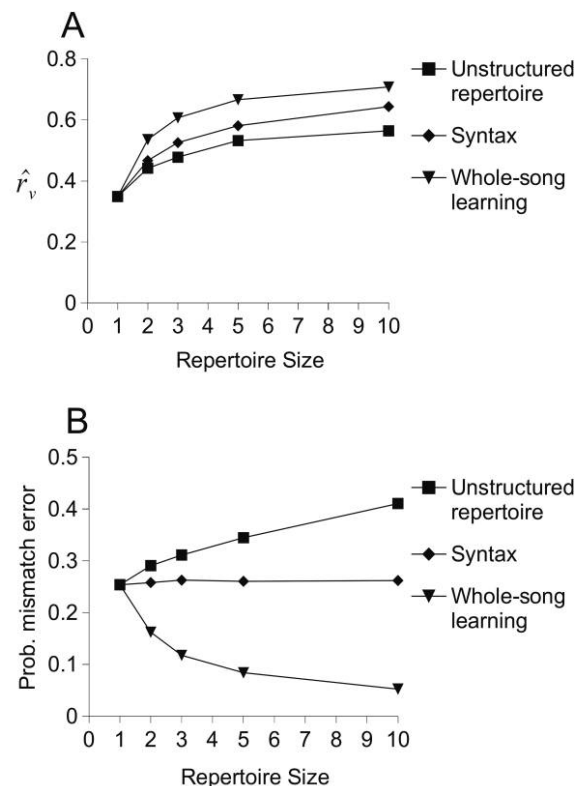


Figure 4: Repertoire complexity and reliability. *A*, Increasing repertoire size increases reliability in simulations of song learning accuracy assessment, but more for whole-song learning than for syntactic learning and more for syntactic learning than for structureless repertoires. *B*, Mismatch error varies with repertoire size for the three styles of repertoire organization. The parameters $d_{i,m}$, $d_{i,p}$ and $d_{s,f}$ were set to 9, and $d_{s,m}$ was set to 1; $M = 1$.

ferent chromosomes are independent). Therefore, the potential for mismatch errors for each element in a syntactically structured repertoire does not increase with repertoire size. Finally, with whole-song learning, mismatch errors are reduced still further. The reason for this is that whole-song learning is functionally equivalent to increasing the dimensionality of the signal. The chance that learning mistakes make two signals converge becomes progressively smaller as the dimensionality of the signals increases: convergence in one dimension is likely to be accompanied by divergence in another dimension.

In summary, the reliability of SLAA is increased if there is a short distance between where males learn songs and where they sing them. When this distance increases to even moderate levels, reliability quickly decreases. In the real world, the only plausible scenario for such a short distance between tutor and tutee is when males learn their songs after dispersing to the area around their territory, a phenomenon that occurs in around 60% of species studied to date (Podos and Warren 2007). Reliability is hampered by a trade-off between novel-song errors, which decrease as females learn over a wider geographical range, and model-as-subject and mismatch errors, which increase at the same time. Mismatch errors can be further reduced, and reliability increased further, by adding structure to how repertoires are organized and learned.

Model 2: Inference of Song Learning Accuracy by Matching to Categories

In model 1, reliability in song learning assessment was limited by confusion between song types and lack of familiarity with potential tutor songs. If females learn enough songs to be familiar with all local tutors, they increase the risk of mismatching a male's song to a similar but incorrect tutor song. The fact that song is culturally transmitted means that there tend to be several similar local variants of a song, which heightens the possibility for confusion. In model 2, we investigate the way in which categorization of song types might reduce such confusion.

The way birds perceive local clusters of songs or song elements has been the subject of experimental research, which has found that "song types" or "note types" form natural categories that are meaningful to the birds themselves (Nelson and Marler 1989; Beecher et al. 1994a; Searcy et al. 1999; Prather et al. 2009). Swamp sparrows, for example, have been shown to distinguish between two variable note types in a categorical fashion on the basis of note length, a continuous variable (Nelson and Marler 1989). More recent work on this system indicates that note-type boundaries differ between closely related populations, suggesting that learning may underlie category development (Prather et al. 2009). Similarly, song sparrows

perceptually lump similar song types into categories in spite of considerable within-category variation (Beecher et al. 1994a; Searcy et al. 1999). There is little evidence as yet whether females cluster songs in the same way as males. Although the details of how note-type and song-type categories develop are poorly understood, it would seem that if learning is involved, birds must use some neural equivalent of a clustering algorithm to determine category boundaries.

In this model, we explore how categorization of song types affects the reliability of SLAA. We assume that learned categories may shape the production and perception of songs and, specifically, that males and females will preferentially select songs to learn near the center of song-type clusters, that is, songs that are more prototypical (fig. 2). Males will therefore learn at most only one song type from a given category, and females will develop one preference for each song type category.

Methods

Model 2 was adapted from model 1. The only change was a categorization step added after tutor songs were sampled: essentially, a basic clustering algorithm. Males and females sample all songs within their song dispersal distance in a random order. The first song to be sampled is placed in a category by itself. The next song is added to the same category if the distance between it and the first song is less than a threshold, s . If not, it becomes the first member of a second category. This process is then repeated for each of the remaining songs in the sample. After a song is added to an existing category, the mean song value is recalculated and then used to represent the category in subsequent steps of the algorithm.

This algorithm is simple enough to make evolutionary simulations practical, but it is not remarkable for the quality of the clusters it produces, because the order in which songs are fed into it influences the category boundaries that are formed. On the other hand, this process may be biologically realistic. It is likely that perceptual categories are formed upon some initial exposure to a novel type and then develop gradually as individuals are exposed to more exemplars.

When males select a song to produce, instead of choosing from a list of learned songs (as in model 1), they choose from the list of category means (if males sing more than one song, then a given category can be represented only once in a male's repertoire). The probability of selecting a given category is the relative proportion of tutor songs belonging to that category within the list of songs the male originally sampled. Similarly, females assess song learning accuracy on basis of the distance between a potential mate's song and the category means that she has learned.

For our default conditions, we set the number of songs in each male's repertoire to 3 and s to 0.02 and assume a structureless repertoire. We set $M = 1$, but because females' preferences are based on means of samples of songs, they generally do not exactly match any particular tutor song. Thus, model-as-subject errors (which can arise when $M < 1$) were much less important in this model than in model 1.

Results

Categorization improved the reliability of SLAA considerably (fig. 5). It also changed the relationship between reliability and the number of songs sampled. In model 1, without categorization, increasing the number of songs sampled (by extending the geographic range of learning or by increasing repertoire size) generally reduced reliability or led to relatively small increases in reliability due to increases in mismatch errors. With categorization, reliability increased as males sampled more songs, by learning either from a broader geographic area (fig. 5A) or from tutors with larger repertoires (fig. 5B). The reason for this difference is that as more songs were sampled, there were more exemplars for each category type. In turn, this led to less variation in how individuals estimated category means and therefore improved reliability.

This principle—that categorization is more effective at improving the reliability of SLAA if more exemplars are sampled per cluster—also applies to the clustering process itself (fig. 5C): the smaller the clustering threshold, s , the narrower the breadth of categories in signal space. Consequently, smaller clustering thresholds led to more clusters per individual, with fewer members of each cluster, and, as predicted, this led to reduced levels of reliability.

In summary, categorization considerably improved the reliability of SLAA. Unlike model 1, this effect was clearest when males and females sampled many different potential tutor songs.

Discussion

Assessment signaling can evolve in the face of uncertainty, as long as signals maintain some reliability on average (Johnstone and Grafen 1993; Kokko 1997; Searcy and Nowicki 2005). In our simulations, we identify two situations in which SLAA can be reasonably reliable: when males learn from within a short distance of their territory and when males and females categorize song types. Any assessment signaling system is likely to be less than completely reliable: variation in signal production induced by factors unrelated to what receivers are assessing, as well as perceptual biases and cognitive limitations on the part of receivers and even the transmission properties of the

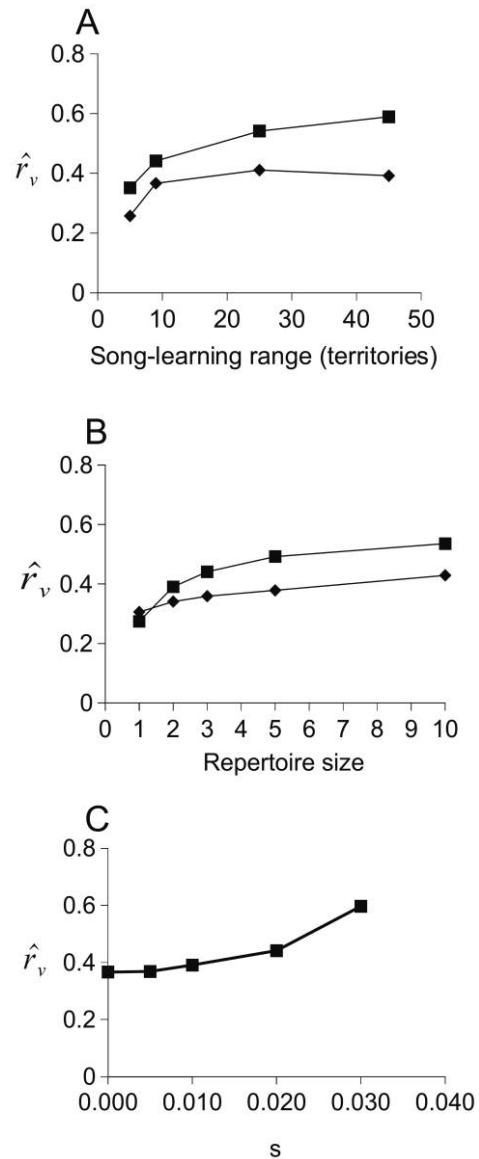


Figure 5: Categorical perception increases the reliability of song learning accuracy assessment over a range of conditions. Squares represent simulations with categorization and diamonds simulations without categorization. A, Varying song learning ranges (the song learning ranges were equal for both males and females). B, Varying repertoire size. C, Varying the threshold for placing a song within a category, s ($s = 0$ means that all songs were placed in separate categories, as in the noncategorization models). Unless otherwise stated, song learning range was set at 9 territories, repertoire size at 3, and s at 0.02.

environment, all act to prevent receivers from assessing signalers perfectly. But our models demonstrate that uncertainty about the identity of the model from which a male learns adds an additional challenge to maintaining reliability unique to song learning accuracy assessment

(SLAA). This additional source of unreliability is present to some degree in all of our simulations, but it does not mean that SLAA cannot evolve: first, females can assess males on the basis of more than one signal (unlike the simplifying assumption we made in our models), and selection may favor females that augment their assessment of males with SLAA; second, as pointed out by Nowicki et al. (1998), SLAA might provide specific information about a certain period of a male's life that is not provided by other signals.

The first scenario under which we found SLAA to be reliable was when males learned their songs at a short distance from their territories (either because they learned after dispersal or because dispersal distances were very short). This situation aids reliability by increasing the probability that a female has encountered the tutor from which a male learned its song. This scenario would be most likely to happen if males learned songs after acquiring a territory, a phenomenon that occurs in many but not all songbird species (Podos and Warren 2007; Catchpole and Slater 2008). Even then, our models found that as song dispersal distances increased to even moderately short distances, reliability rapidly decreased. One reason for this result is that the number of potential tutors a female should learn from to maximize reliability increases as a square function of song dispersal distance, and the more songs a female has learned, the greater the probability of mismatch errors (fig. 1B). In general, this scenario is also highly dependent on the cognitive abilities of female receivers. In our models, we did not limit females' abilities to memorize or compare songs, partly because of the lack of empirical evidence to make realistic estimates, but it seems likely that this scenario would make unrealistically large cognitive and time demands on females if males learned from tutors more than a few territories away from their own territory or had complex repertoires. A second limitation of this scenario is how females deal with reencountering males they have already learned from (model-as-subject error). One factor in this is the timing of female learning, which remains largely unknown (Riebel 2009). If females learn only early in life (e.g., Anderson 2009), as in our models, then there is a relatively high probability that they will mistake a former tutor for a younger male that has accurately learned his song. But if, for example, they learn anew at the beginning of each breeding season, then they may be able to avoid this error. In summary, this scenario allows SLAA to be reliable, but it may apply to only very few species.

The second scenario increases the reliability of SLAA by relying on categories of song types rather than specific exemplars, thereby sidestepping the errors in inferring the tutor song. In our models of categorization, individuals learn abstract prototypical versions of song categories

rather than particular instances of songs. This increases the concordance between males and females about the target for learning, because as more songs are sampled, the prototypes of different birds tend to converge. Thus, a female may form an accurate concept of what a male is trying to reproduce, even if she does not share the same tutors (fig. 4). Categorization thus removes the problem of novel-song error. There is experimental evidence that birds do categorize both song types (e.g., Beecher et al. 1994a; Searcy et al. 1999) and the subunits of songs (Nelson and Marler 1989; Prather et al. 2009), but it is not yet known whether birds form prototypes of these categories or whether they use such prototypes as models for song learning. A recent study of swamp sparrows (R. F. Lachlan, R. C. Anderson, and S. Nowicki, unpublished data) found that females preferred songs that were more prototypical examples of population-wide song-type categories over ones that were outliers, providing the first piece of evidence in support of this hypothesis.

In both scenarios, we found that increasing repertoire complexity increased the reliability of SLAA. This was because larger repertoires provided more information to females, either about the particular male she was assessing (both scenarios) or about the nature of population-wide categories (scenario 2 only). We believe this is the first time a positive relationship has been suggested between these two facets of song learning: the amount an individual learns and the quality with which it learns. In addition to this, there may be a trade-off between the two if the additional cognitive demands of learning larger repertoires decrease the accuracy of learning or limit the amount of songs learned by males or females.

As with all models, we made assumptions that influenced our results. First, we used a simple model of sexual selection in which male condition influenced parental ability but was not itself inherited. We did this because it made our models more computationally tractable, but in general we would expect very similar results if male condition is partly determined by a large number of underlying genes. One corollary of the approach we took is that there was no extrapair paternity in our model, even though this phenomenon is widespread in songbirds (Griffith et al. 2002). Extrapair paternity would be expected to influence some aspects of our model by skewing reproductive success in males. In particular, we would expect that if older males are available for extrapair copulations, model-as-subject errors might be even more important in reality than in our models.

Second, in our models we assumed no limits in the precision of female perception or the accuracy of their memory: females could detect any differences between males, however small. Although this assumption would, in general, affect any type of signal assessment, in our

model it particularly increased the rate of model-as-subject errors. Because males did not change their songs over time (also an assumption), a male that had served as a female's tutor earlier in life would have the maximum level of attractiveness to her if encountered later as a potential mate. In the real world, it seems plausible that females might not be able to distinguish between an older male's song and a particularly accurately learned copy of it sung by a younger male, although we know of no empirical evidence addressing this issue.

Third, our models simplified songs to just a few continuous dimensions in which song culturally evolved freely without any selection. But bird song is a complex signal, often with multiple functions (including territorial defense as well as mate attraction), combining several routes for communication (such as preferences for more complex songs or for higher-performance songs), that may be produced by females as well as males (see Catchpole and Slater 2008 for an overview). For the sake of the tractability of our models, we did not consider these additional aspects of bird song communication. In some cases, other hypotheses that explain the evolution of accurate learning might complement our models of SLAA. Hypotheses in which accurate learning allows females to select local males over foreign males no doubt involve females categorizing local song types or other features of local songs, for example. It is especially worth considering that in some species, song might not be involved in the particular interaction central to our models: the assessment of potential mates by females. On the other hand, we think it plausible that SLAA might be employed in other types of assessment, such as territorial interactions between males. In line with this, a recent experiment demonstrated a stronger territorial response by male swamp sparrows to songs that were more prototypical of their song type than to songs that were outliers, in line with the predictions of our categorical model (R. F. Lachlan, R. C. Anderson, and S. Nowicki, unpublished data).

Natural communication systems in which different socially learned signals communicate different information are unusual: aside from songbird song, human language (and perhaps other aspects of human culture) and bottlenose dolphin signature whistles (e.g., Janik and Slater 1998) are the only well-documented examples. The problem we address in this article—how a “meaning” can be extracted from a signal in the face of inaccurate learning and dynamic cultural evolution—is common to all of them. In SLAA, the “meaning” of a signal has something to do with mate assessment, and this assessment is completely accurate only if a male's signal, his song, can be matched to the specific model that he learned.

SLAA is not the only hypothesis that might explain why songs are learned accurately, but in other hypotheses, as-

essed songs need only be matched to broader song-type categories, in the context of either sharing between neighbors (e.g., Vehrencamp 2000; Lachlan et al. 2004; Beecher and Brenowitz 2005) or producing songs found in the local population (e.g., Nottebohm 1969; Baker 1982). Likewise, in a different communication system, human speech, a sound need only be matched to the appropriate word category. Instead, SLAA makes the clear prediction that receivers should distinguish between exemplars that belong to the same song-type category. Nevertheless, we think that two of the types of error we documented here—novel-song error and mismatch error—may be relevant to all of these forms of communication. For example, it is essentially a mismatch error when a British English speaker hears the word “pin” when a Southern US English speaker utters “pen,” and when a British speaker does not recognize the word “chitlins” uttered by a Southern US speaker, that is similar to a novel-song error in our model. It seems plausible that some of the solutions to these problems we modeled here, such as increasing the complexity of the signal and relying on categorization to shape learning, might apply to other systems, too, and might reflect deep similarities found between learned communication systems.

Acknowledgments

We thank E. Derryberry, D. Irwin, I. Liu, S. Peters, and two anonymous reviewers, who provided useful comments on the manuscript. Funding was provided by the Duke University Provost Office.

Literature Cited

- Anderson, R. C. 2009. Operant conditioning and copulation solicitation display assays reveal a stable preference for local song by female swamp sparrows *Melospiza georgiana*. *Behavioral Ecology and Sociobiology* 64:215–223.
- Baker, M. C. 1982. Vocal dialect recognition and population genetic consequences. *American Zoologist* 22:561–569.
- Beecher, M. D., and E. A. Brenowitz. 2005. Functional aspects of song learning in birds. *Trends in Ecology & Evolution* 20:143–149.
- Beecher, M. D., S. E. Campbell, and J. M. Burt. 1994a. Song perception in the song sparrow: birds classify by song type but not by singer. *Animal Behaviour* 47:1343–1351.
- Beecher, M. D., S. E. Campbell, and P. K. Stoddard. 1994b. Correlation of song learning and territory establishment strategies in the song sparrow. *Proceedings of the National Academy of Sciences of the USA* 91:1450–1454.
- Brooks, R., and D. J. Kemp. 2001. Can older males deliver the good genes? *Trends in Ecology & Evolution* 16:308–313.
- Brumm, H., S. A. Zollinger, and P. J. B. Slater. 2009. Developmental stress affects song learning but not song complexity and vocal

- amplitude in zebra finches. *Behavioral Ecology and Sociobiology* 63:1387–1395.
- Buchanan, K. L., S. Leitner, K. A. Spencer, A. R. Goldsmith, and C. K. Catchpole. 2004. Developmental stress selectively affects the song control nucleus HVC in the zebra finch. *Proceedings of the Royal Society B: Biological Sciences* 271:2381–2386.
- Catchpole, C. K., and P. J. B. Slater. 2008. *Bird song: biological themes and variations*. 2nd ed. Cambridge University Press, Cambridge.
- Gil, D., M. Naguib, K. Riebel, A. Rutstein, and M. Gahr. 2006. Early condition, song learning and the volume of song brain nuclei in the zebra finch (*Taeniopygia guttata*). *Journal of Neurobiology* 66:1602–1612.
- Grant, B. R., and P. R. Grant. 1996. Cultural inheritance of song and its role in the evolution of Darwin's finches. *Evolution* 50:2471–2487.
- Grieser, D., and P. K. Kuhl. 1989. Categorization of speech by infants: support for speech-sound prototypes. *Developmental Psychobiology* 25:577–588.
- Griffith, S. C., I. P. F. Owens, and K. A. Thuman. 2002. Extra pair paternity in birds: a review of interspecific variation and adaptive function. *Molecular Ecology* 11:2195–2212.
- Holveck, M.-J., A. C. Vieira de Castro, R. F. Lachlan, C. ten Cate, and K. Riebel. 2008. Accuracy of song syntax learning and singing consistency signal early condition in zebra finches. *Behavioral Ecology* 19:1267–1281.
- Hughes, M., S. Nowicki, W. A. Searcy, and S. Peters. 1998. Song-type sharing in song sparrows: implications for repertoire function and song learning. *Behavioral Ecology and Sociobiology* 42:437–446.
- Janik, V. M., and P. J. B. Slater. 1998. Context-specific use suggests that bottlenose dolphin signature whistles are cohesion calls. *Animal Behaviour* 56:829–838.
- Johnstone, R. A. 1996. Multiple displays in animal communication: “backup signals” and “multiple messages.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 351:329–338.
- Johnstone, R. A., and A. Grafen. 1993. Dishonesty and the handicap principle. *Animal Behaviour* 46:759–764.
- Kirkpatrick, M., and Ryan, M. J. 1991. The evolution of mating preferences and the paradox of the lek. *Nature* 350:33–38.
- Kluender, K., A. J. Lotto, L. L. Holt, and S. L. Bloedel. 1998. Role of experience for language-specific functional mappings of vowel sounds. *Journal of the Acoustical Society of America* 104:3568–3582.
- Kokko, H. 1997. Evolutionarily stable strategies of age-dependent sexual advertisement. *Behavioral Ecology and Sociobiology* 41:99–107.
- Lachlan, R. F., V. M. Janik, and P. J. B. Slater. 2004. The evolution of conformity-enforcing behaviour in cultural communication systems. *Animal Behaviour* 68:561–570.
- Lachlan, R. F., and P. J. B. Slater. 2003. Song learning in chaffinches: how accurate, and from where? *Animal Behaviour* 65:957–969.
- Lachlan, R. F., L. Verhagen, S. Peters, and C. ten Cate. 2010. Are there species-universal categories in bird song phonology and syntax? a comparative study of chaffinches (*Fringilla coelebs*), zebra finches (*Taeniopygia guttata*), and swamp sparrows (*Melospiza georgiana*). *Journal of Comparative Psychology* 124:92–108.
- Lahti, D. C., D. L. Moseley, and J. Podos. 2011. A tradeoff between performance and accuracy in bird song learning. *Ethology* 117:1–10.
- Lynch, A. 1996. The population memetics of birdsong. Pages 181–197 in D. E. Kroodsma and E. H. Miller, eds. *Ecology and evolution of animal communication*. Cornell University Press, Ithaca, NY.
- MacDonald, I. F., B. Kempster, L. Zanette, and S. A. MacDougall-Shackleton. 2006. Early nutritional stress impairs development of a song-control brain region in both male and female juvenile song sparrows (*Melospiza melodia*) at the onset of song learning. *Proceedings of the Royal Society B: Biological Sciences* 273:2559–2563.
- Marler, P., and S. Peters. 1981. Sparrows learn adult song and more from memory. *Science* 213:780–782.
- Nelson, D. A., and P. Marler. 1989. Categorical perception of a natural stimulus continuum: birdsong. *Science* 244:976–978.
- Nottebohm, F. 1969. The origins of vocal learning. *American Naturalist* 106:116–140.
- Nowicki, S., S. Peters, and J. Podos. 1998. Song learning, early nutrition and sexual selection in songbirds. *American Zoologist* 38:179–190.
- Nowicki, S., W. A. Searcy, and S. Peters. 2002a. Brain development, song learning and mate choice in birds: a review and experimental test of the “nutritional stress hypothesis.” *Journal of Comparative Physiology A* 188:103–114.
- . 2002b. Quality of song learning affects female response to male bird song. *Proceedings of the Royal Society B: Biological Sciences* 269:1949–1954.
- Podos, J., and P. S. Warren. 2007. The evolution of geographic variation in birdsong. *Annual Review of Ecology and Systematics* 37:403–458.
- Prather, J. F., S. Nowicki, R. C. Anderson, S. Peters, and R. Mooney. 2009. Categorical perception in learned vocal communication. *Nature Neuroscience* 12:221–228.
- Riebel, K. 2009. Song and female mate choice in zebra finches: a review. *Advances in the Study of Behavior* 40:197–238.
- Ritchie, G., S. Kirby, and D. Hawkey. 2008. Song learning as an indicator mechanism: modelling the developmental stress hypothesis. *Journal of Theoretical Biology* 251:570–583.
- Searcy, W. A., and S. Nowicki. 2005. *The evolution of animal communication: reliability and deception in animal signaling systems*. Princeton University Press, Princeton, NJ.
- Searcy, W. A., S. Nowicki, and S. Peters. 1999. Song types as fundamental units in vocal repertoires. *Animal Behaviour* 58:37–44.
- Soha, J. A., and P. Marler. 2001. Vocal syntax development in the white-crowned sparrow (*Zonotrichia leucophrys*). *Journal of Comparative Psychology* 115:172–180.
- Tchernichovski, O., P. P. Mitra, T. Lints, and F. Nottebohm. 2001. Dynamics of the vocal imitation process: how a zebra finch learns its song. *Science* 291:2564–2569.
- Thorpe, W. H. 1958. The learning of song patterns by birds, with especial reference to the song of the chaffinch, *Fringilla coelebs*. *Ibis* 100:535–570.
- Vehrencamp, S. L. 2000. Is song-type matching a conventional signal of aggressive intentions? *Proceedings of the Royal Society B: Biological Sciences* 268:1637–1642.
- Zann, R., and E. Cash. 2008. Developmental stress impairs song complexity but not learning accuracy in non-domesticated zebra finches (*Taeniopygia guttata*). *Behavioral Ecology and Sociobiology* 62:391–400.

Associate Editor: Tony D. Williams
 Editor: Judith L. Bronstein