Anim Cogn (2017) 20:309–320
DOI 10.1007/s10071-016-1053-7

ORIGINAL PAPER

Song learning and cognitive ability are not consistently related in a songbird

Rindy C. Anderson[1] · **William A. Searcy**[2] · **Susan Peters**[3] · **Melissa Hughes**[4] · **Adrienne L. DuBois**[2] · **Stephen Nowicki**[3]

Received: 8 February 2016 / Revised: 10 October 2016 / Accepted: 3 November 2016 / Published online: 14 November 2016

Abstract Learned aspects of song have been hypothesized to signal cognitive ability in songbirds. We tested this hypothesis in hand-reared song sparrows (*Melospiza melodia*) that were tutored with playback of adult songs during the critical period for song learning. The songs developed by the 19 male subjects were compared to the model songs to produce two measures of song learning: the proportion of notes copied from models and the average spectrogram cross-correlation between copied notes and model notes. Song repertoire size, which reflects song complexity, was also measured. At 1 year of age, subjects were given a battery of five cognitive tests that measured speed of learning in the context of a novel foraging task, color association, color reversal, detour-reaching, and spatial learning. Bivariate correlations between the three song measures and the five cognitive measures revealed no significant associations. As in other studies of avian cognition, different cognitive measures were for the most part not correlated with each other, and this result remained true when 22 hand-reared female song sparrows were added to the analysis. General linear mixed models controlling for effects of neophobia and nest of origin indicated that all three song measures were associated with better performance on color reversal and spatial learning but were associated with worse performance on novel foraging and detour-reaching. Overall, the results do not support the hypothesis that learned aspects of song signal cognitive ability.

Keywords Song learning · Cognition · Song complexity · Songbird · Song sparrow · *Melospiza melodia*

Introduction

Sexual selection acting through mate choice has been suggested to play a role in the evolution of increased cognitive ability (Jacobs 1996; Miller 2000; Boogert et al. 2011b; Peters et al. 2014). This hypothesis is supported by evidence that females prefer mates with better cognitive ability in a number of animals, including fish (Shohet and Watt 2009), birds (Keagy et al. 2009, 2011), rodents (Spritzer et al. 2005), and humans (Li et al. 2002; Prokosch et al. 2009). In many cases, however, it is not clear how females could assess the cognitive abilities of potential mates (Boogert et al. 2011b), which is a necessary first step if mates with better cognition are to be preferred. Such assessment would be greatly facilitated if males advertised their cognitive ability via their displays, but cognition has no obvious link to many common sexual displays, such as bright colors, exaggerated tails and plumes, pheromones, and so forth. Sexual displays that are learned behaviors constitute an exception, where a link to cognition seems more plausible (Peters et al. 2014). The best known case of a learned sexual display is the song of songbirds. Here we explore the possibility that learned aspects of song signal cognitive ability in one species of songbird, the song sparrow (*Melospiza melodia*).

✉ William A. Searcy
wsearcy@miami.edu

[1] Department of Biological Sciences, Florida Atlantic University, Boca Raton, FL, USA

[2] Department of Biology, University of Miami, Coral Gables, FL, USA

[3] Department of Biology, Duke University, Durham, NC, USA

[4] Department of Biology, College of Charleston, Charleston, SC, USA

Several aspects of the biology of songbirds accord with the idea that songs might signal cognitive ability in this group (Nowicki et al. 2000; Boogert et al. 2008; Peters et al. 2014). Because song development is affected by learning in songbirds (Marler and Mundinger 1971; Beecher and Brenowitz 2005), song qualifies as a cognitive trait by usual definitions of cognition (Shettleworth 1998, 2010). In some animals, different cognitive abilities are consistently positively associated with each other (Spearman 1904; Anderson 1993; Plomin 2001; Matzel et al. 2003); if this holds true in songbirds, then song learning ability ought to be positively associated with performance on other, more typical measures of cognition. Developmental stresses that are known to cause deficits in song learning, such as nutritional limitation (Nowicki et al. 2002a; Spencer and MacDougall-Shackleton 2011), also have a negative impact on other aspects of cognitive performance (Pravosudov et al. 2005; Farrell et al. 2016) and thus should contribute to an association between these behaviors. Genetic problems such as inbreeding are likely to impair both song (Reid et al. 2005) and cognition in general (Bashi 1977), again potentially contributing to a positive association between song and other cognitive measures. Despite these strong, a priori arguments for associations between learned aspects of song and cognitive performance, results of previous studies on such associations have been mixed (Peters et al. 2014). Here we test for associations between song and performance on cognitive tests in male song sparrows that have been hand-reared and tutored with recorded song, allowing us to make more precise measurements of success in song learning than have been possible in previous studies.

Previous studies relating cognitive ability to song have focused solely on measures of song complexity. In the first such study, Boogert et al. (2008) found that the success of male zebra finches (*Taeniopygia guttata*) in learning a novel foraging skill was positively associated with number of elements per song but showed no association with the number of unique elements per song or with song duration. In a second zebra finch study, in which birds were tested in flocks rather than singly, Templeton et al. (2014) found that no measure of song complexity was associated with success in acquiring the same novel foraging skill or with success on two other tasks. In two studies of song sparrows, song repertoire size was positively associated with success in a detour-reaching task (Boogert et al. 2011a), was negatively associated with success in color association reversal (Boogert et al. 2011a) and spatial learning (Sewall et al. 2013), and showed no association with success in novel foraging and color association (Boogert et al. 2011a). Finally, in a study of European starlings (*Sturnus vulgaris*), song bout length correlated positively with performance on a spatial learning task, but not with performance on a social learning task (Farrell et al. 2012). In sum, measures of song complexity in previous studies have often shown no correlation with cognitive measures and have been about as likely to show negative correlations as positive ones.

Song complexity is a logical feature to examine in studies relating cognition to how males sing, as there is evidence in each of the study species for female preferences based on the focal measure of complexity: for greater syllable diversity in zebra finches (Riebel 2009; Vyas et al. 2009), for larger song repertoire size in song sparrows (Searcy 1984; Reid et al. 2004), and for greater song bout length in starlings (Eens et al. 1991). The fact that females exercise preferences for these song features strengthens the possibility that the features advertise aspects of male quality such as cognitive ability. Song complexity may not, however, reflect song learning ability as directly as do other potential song measures. The most direct way to assess the quality of song learning is to compare the songs produced by young males to the model songs they have heard during their critical learning periods. Accurate copying of models then equates unambiguously with superior song learning.

In the wild, young song sparrows copy songs that they hear produced by territorial males in the neighborhoods in which they will later establish their own territories (Beecher et al. 1994; Nordby et al. 1999). In captivity, learning of recorded models is concentrated in a sensitive period spanning 20–60 days post-hatching, though some learning can occur substantially later (Marler and Peters 1987). In the eastern populations that we study, song sparrows in both the laboratory and field develop songs that combine elements copied from model songs with elements that cannot be traced to any model (Marler and Peters 1988; Hughes et al. 1998). Nowicki et al. (2002b) showed that when responding to the songs of hand-reared males tutored with recorded song, wild-caught female song sparrows give stronger courtship responses to songs that are good copies of songs from the source population than to songs that are less well copied. In the present study, we assess song learning using the same two measures associated with female response in the Nowicki et al. (2002b) study: the proportion of notes per song type that are copied from model songs, and the average accuracy of copying of those notes as measured by spectrogram cross-correlation. For comparison with other song sparrow studies (Boogert et al. 2011a; Sewall et al. 2013), we use song repertoire size as a third measure of song learning.

While two of the three song measures we use are new to studies of song and cognition, all five of our cognitive measures involve tasks similar to ones featured in previous studies. The five tasks are: a novel foraging task, previously used by Boogert et al. (2008, 2011a); color association and color reversal tests, previously used by Boogert et al. (2011a); a spatial learning task similar to one used by Sewall et al. (2013); and a detour-reaching task, used by Boogert et al. (2011a). We use these five tasks to test the hypothesis that quality of song learning is associated with general cognitive ability. The prediction we test is that speed of learning of some or all of these five tasks will be positively associated with our three measures of song learning in male song sparrows. We also tested the assumption that different cognitive abilities are positively associated by measuring correlations in performance among the five cognitive tasks. For these tests, we increased our sample size by adding data on hand-reared female song sparrows to our data on hand-reared males.

Methods

Hand-rearing and song tutoring

The subjects were 19 male and 22 female song sparrows collected from nests in Crawford County, Pennsylvania, USA, during May of 2013. Birds were collected at 3–6 days post-hatching and transported overnight to a laboratory at Duke University, where they were hand-reared. Birds were initially hand-fed a standard diet (Marler and Peters 1988), first at 30-min intervals and later at 60-min intervals. Starting at 15 days, the birds were given access to seeds and other food for self-feeding; we then gradually reduced hand-feeding until the birds were totally self-sufficient at about 24 days. The young birds were housed socially except on days when they were recorded.

All subjects were tutored with recorded songs starting at approximately 10 days post-hatching and continuing for 12 weeks, thus spanning the entire sensitive period for song learning (Marler and Peters 1987). For tutoring, we used single variants of 32 songs recorded during 2009 and 2010 from male song sparrows in the same Pennsylvania sites where the subjects were collected. Tutor songs were recorded with a Marantz PMD 660 digital recorder, a Shure SM57 microphone, and a Sony 330 parabolic reflector. The 32 songs were chosen out of a much larger sample based on clarity of recording and on distinctiveness from each other. Distinctive model songs were used to aid in matching learned songs back to specific models. All model song types were presented during two 2 ½ h sessions each day, one in the morning and one in the afternoon. During each session, each tutor song was played at a rate of 6 songs per minute for a 4-min period, with 1 min of silence between bouts. The song bout order was randomized each week. Songs were played at 75–80 dB (at 30 cm) from an Advent Powered Partner AV570 speaker to subjects that were 0.6–2.0 m distant.

Song analysis

We recorded complete song type repertoires from the male subjects once they completed song development. In song sparrows, song is quite rare in females in nature (Arcese et al. 1988) and appears to be even rarer in hand-reared females (S. Peters pers. obs.), so female subjects were not included in analyses relating song learning to cognition. In wild song sparrows, 200 songs are usually sufficient to capture complete male repertoires, with an additional song type sometimes being found between 200 and 300 songs (Searcy et al. 1985). Here we recorded an average of 1550 songs per male (range 254–2907) over a 5-week period beginning at about 11 months post-hatching. Spectrograms were prepared of the subjects' song types and the tutor song types using Syrinx (J. M. Burt, www.syrinxpc.com) (512 transform, 10 ms/line time axis, 0–10,000 frequency axis) and Signal for Windows v. 4 (Engineering Design, Berkeley CA) (172.3 Hz frequency resolution and 5.8 ms time resolution). Song sparrows from our study population have repertoires averaging about 8 song types, which are sung with considerable within-type variation (Peters et al. 2000). We identified all the variants for each song type we recorded and determined which was the most common. Subsequent analysis concentrated on these most common variants, which on average constituted 44% (±11% SD) of a male's songs. Another 4% of songs on average consisted of the same notes as the most common variant in a different order, giving a total of 48% (±13%) of songs that would give identical results on our two song learning measures (see below). The next most common variant differing in one or more notes from the most common variant was on average substantially less common (18 ± 3%).

Two experienced observers independently compared the most common variant of each song type with the set of 32 tutor songs to identify which of the subjects' notes and phrases were copied from tutor songs and which tutor songs they were copied from. Initial agreement on which model note or phrase that each subject note or phrase had been copied from was 78% (±13% SD) per subject. Cases of disagreement about copying were settled by consensus. From these data, we calculated for each subject our first measure of quality of song learning: the average proportion of notes per song type copied from tutor songs (proportion copied). For all notes identified as copied (about 80% of a male's notes on average), we used Signal for Windows v.4 (transform length 256, time increment 1.0 ms) to calculate the spectrogram cross-correlation between the subject's copy and the original note in the tutor song. The average cross-correlation per male for all copied notes was our second measure of song learning quality.

Cognitive tests

Female subjects were administered cognitive tests starting in March 2014 when they were approximately 10 months old. Male subjects were tested starting in June of 2014 when they were approximately 13 months old. In both sets of tests, subjects were presented first with a test for neophobia and then with five cognitive tasks in a set order: (1) novel foraging, (2) color association, (3) color reversal, (4) spatial learning, and (5) detour-reaching. A set order of tasks was used because our intent was not to test whether subjects performed better on one task than another but rather to correlate performance on each task with song measures. Subjects were tested in the afternoon after 5 h of food deprivation, imposed to ensure that the birds were motivated to work for food rewards. For all tasks, the food rewards were larvae of the mealworm beetle, *Tenebrio molitor*, a preferred food item. Continued motivation to eat was tested each day after completion of the cognitive tasks. The average latency to feed at the end of testing sessions was less than 30 s, confirming high motivation. Trials for all tests were run on one subject at a time in its home cage, with the results viewed remotely via video. The birds were visually isolated from each other during testing.

Four of the cognitive tasks (novel foraging, color association, color reversal, and detour-reaching) were adapted from Boogert et al. (2011a), and the fifth (spatial learning) was more substantially modified from Sewall et al. (2013). We describe the neophobia test and the five cognitive tests below. For all tasks, we define a criterion for successful learning, and the summary measure of success is the number of trials taken to reach that criterion.

Neophobia

Neophobia was measured as the time taken by a subject to remove a mealworm from a foraging grid after the grid was first placed in its cage. Foraging grids were 13.5 × 9 × 2.5 cm blocks of plastic (Fig. 1a) that were also used in the first three cognitive tasks (see below).

Novel foraging

Each foraging grid contained 6 wells 1.3 cm in diameter and 0.8 cm deep. Four wells were baited with a mealworm and then covered with a plastic lid that fit snugly over the well, hiding the mealworm. For the novel foraging task, subjects had to learn to remove the lid to obtain the mealworm. Almost all subjects did this by grasping the edge of the lid with their bill and flipping it away. Song sparrows are not known to perform a similar motion in natural foraging. Subjects learned this task in five stages: (1) Mealworms were placed in four wells with no lids; (2) lids were positioned adjacent to the four baited wells; (3) lids half-covered the baited wells; (4) lids were placed to cover loosely the baited wells; and (5) lids were fitted snugly over the baited wells. The criterion for a successful trial was to obtain at least two of the available mealworms within 2 min. Subjects had to succeed on three of four consecutive trials to pass stages 2–4 and four of five consecutive trials to pass stage 5.

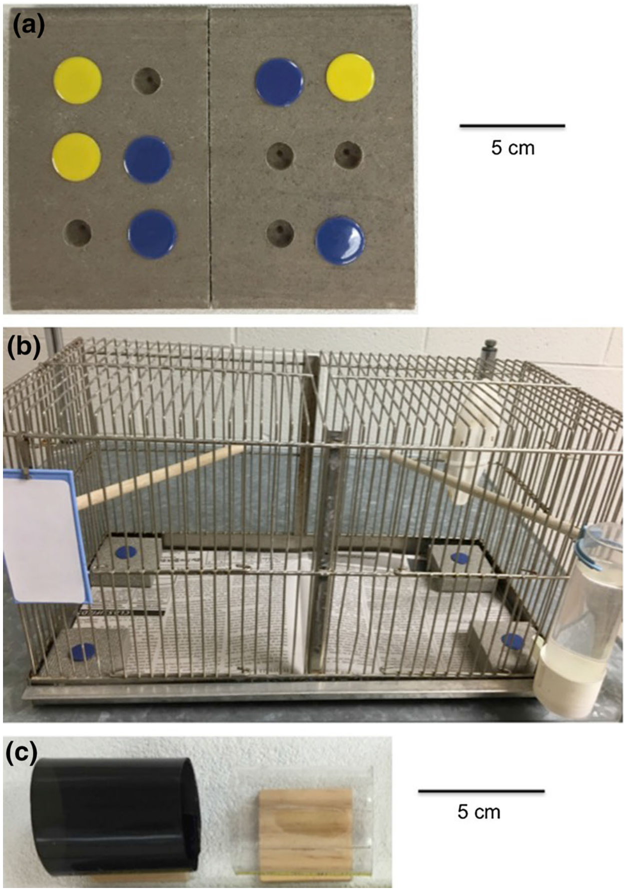

Fig. 1 Photographs of the apparatus used in the cognitive tests. **a** Two of the foraging grids used in the novel foraging, color association, and color reversal tasks. Mealworms are placed in some of the wells and covered with one of the *blue* or *yellow lids*. A 5-cm line is provided for scale. **b** A cage set up for the spatial learning task. Blocks with a single well are placed in each corner of the cage. The block in one corner is consistently baited with a mealworm. The cage bottoms are 46 × 22 cm. **c** The black (*left*) and clear (*right*) tubes used in the detour-reaching task. The *black tube* is used in training. The test requires removing a mealworm from one of the open ends of the clear cylinder without attempting to peck through the walls. A 5-cm line is provided for scale (color figure online)

Color association

Using two six-well grids pushed together, four wells were covered with blue lids and four with yellow lids, with positions shuffled arbitrarily between trials. For each subject, one color was chosen to be rewarded and one to be unrewarded, with color choices balanced across subjects. Then in all trials, one mealworm was placed in each well covered by a lid of the rewarded color, while no reward was placed in wells covered by the alternate color. The criterion for success in this task was to remove all four lids of the rewarded color before removing any lids of the unrewarded color on six of seven consecutive trials.

Color reversal

This task was set up exactly as the color association task, except that the reward scheme was reversed for each subject—that is, the unrewarded color in the previous task now was rewarded and the previously rewarded color was now unrewarded. The criterion for success remained the same: removing all four lids of the rewarded color before any lids of the unrewarded color on six of seven consecutive trials.

Spatial learning

This task employed four small blocks each made from the same material as the six-well grids but containing only a single well. For each trial, one block was placed in each of the four corners of the subject's 46 × 22 × 26 cm cage (Fig. 1b). Wells were covered with lids for all blocks, but only one well was baited with a mealworm. For each subject, one corner was chosen randomly, and the baited block was placed in that corner on all trials. A subject succeeded on a single trial if it removed the lid on the rewarded block first. The criterion for success on the task as a whole was to choose the correct block on six of seven consecutive trials. After a bird passed the spatial task, we conducted a final probe trial to test whether the birds were using direct sensory cues rather than memory to find the baited block. In this probe, we baited a different corner and ran the trial as usual. Birds visited the baited block in only 3% of the probe trials, indicating that they could not sense the reward directly.

Detour-reaching

The goal of this task was to learn to remove a food reward (a mealworm) from a clear plastic cylinder through one of the open ends without attempting to peck at the food through the solid walls of the cylinder. Subjects were first presented with a baited cylinder having opaque black walls (Fig. 1c). Once a subject succeeded in removing the mealworm without pecking at the sides of the cylinder in four of five consecutive trials, it was advanced to the clear cylinder stage (Fig. 1c). The criterion for success on this stage was to remove the mealworm without error (i.e., without pecking the cylinder walls) in six of seven consecutive trials. Success was measured by the number of trials needed to reach criterion on the clear cylinder stage.

Data analysis

Some of the cognitive measures and song measures passed the Shapiro–Wilk test for normality ($P > 0.05$), but others were non-normal and could not be normalized with standard transforms. Accordingly, we rely for the most part on nonparametric statistical tests. The emphasis in the analysis is on examining the association between the song measures and the male cognitive results. The data on females are used only in combination with the male data to test for associations between individual performance on different cognitive tests. With the exception of the GLMMs (below), all the statistical tests were run in SPSS v. 22.

We tested for associations between song measures and cognitive measures using Spearman rank correlations. Associations between different cognitive measures were also assessed with Spearman rank correlations. The various cognitive scores were combined using principal component analysis (PCA), a standard method for combining cognitive measures in both mammals (Plomin 2001; Locurto et al. 2003) and birds (Boogert et al. 2011a; Keagy et al. 2011; Isden et al. 2013). An unrotated solution was used. We also used PCA to combine the three song measures. One criterion for sample size for PCA is that the number of subjects be at least five times the number of variables (O'Rourke and Hatcher 2013); our sample exceeds that criterion for the song measures ($19/3 = 6.3$), but not for the cognitive measures ($19/5 = 3.8$). We used PCA in both analyses to facilitate comparison with other studies of bird song and cognition.

We also tested whether the song traits predict the cognitive measures using a generalized linear mixed model estimated using the lme4 package (Bates et al. 2015) in R (R Core Team 2014). We modeled repeated performance outcomes using a Poisson distribution, including random intercepts by nest of origin, by individual bird (allowing the latter to be a function of each bird's measured neophobia) and by task (to allow for different baseline performances across tasks).

Results

Song learning

Some of the songs produced by the 19 hand-reared males were excellent copies of entire model songs, whereas others of the subjects' songs showed little evidence of having been copied from any model (Fig. 2). The mean proportion of notes copied from model songs, averaged

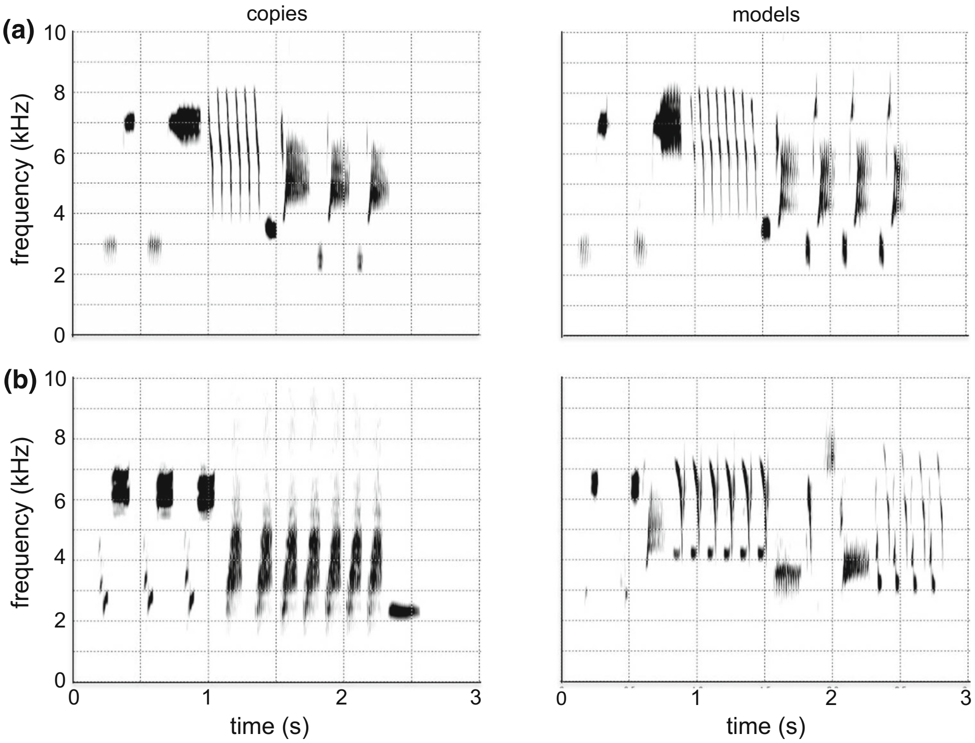

Fig. 2 Spectrograms of two songs sung by hand-reared males ("copies") and the two tutor songs they were copied from ("models"). **a** At *left* is a song from a subject male that is a good copy of the model song at *right*, good both in that a high proportion of notes are copied from the model and in that the mean accuracy of note copying is high. **b** At *left* is a song from a subject male that is a poor copy of the model song on the *right*. In both cases, no other model song was a better match to the copied song

over a subject's song types, varied from 0.461 to 0.994, with an overall mean of 0.799. The mean cross-correlation between subject note and model note, averaged over all copied notes for a subject, ranged from 0.457 to 0.707, with an overall mean of 0.600. Repertoire sizes of the 19 males varied from 4 to 11 song types, with a mean of 6.4. Thus, all three of the song measures showed substantial variability among subjects.

Neophobia

In the tests with male subjects, neophobia was positively correlated with novel foraging ($r_s = 0.487$, $P = 0.034$) and negatively correlated with detour-reaching ($r_s = -0.532$, $P = 0.019$). The correlations between neophobia and the remaining cognitive measures were all not significant ($P > 0.05$). Neophobia is measured on a positive scale (higher scores indicate greater neophobia), whereas all the cognitive traits are measured on an inverse scale (higher scores indicate worse cognition). Thus, the positive correlation between neophobia and novel foraging suggests that neophobia interfered with success on the novel foraging task. In contrast, the negative correlation between neophobia and detour-reaching suggests that greater neophobia was associated with better success on this task.

Song and cognition

Table 1 shows Spearman rank correlations between the three measures of song learning quality (proportion copied, average cross-correlation, and repertoire size) and success on the five cognitive tasks for males. None of the 15 correlations are significant (all $P > 0.05$). The hypothesis that song signals cognitive ability predicts negative correlations, because all three song measures scale positively (higher scores indicate better song learning), while (again) all the cognitive measures are on inverse scales (higher scores indicate worse cognitive performance). Eight of the non-significant correlations were negative, as predicted, and seven were positive, counter to prediction.

Associations between cognitive measures

Table 2a shows Spearman rank correlations between the different cognitive measures within the male tests. Success on color association is significantly, positively correlated with success on color reversal. No other correlations are significant. Table 2b shows the correlations between the different cognitive measures combining the male and female data. In the combined data, performance on color association is again positively correlated with success on color reversal. None of the other correlations is significant.

Table 1 Spearman rank correlations between cognitive measures and song learning measures

Song measures	Novel foraging	Color association	Color reversal	Spatial learning	Detour-reaching
Proportion copied	0.026	−0.112	−0.210	−0.067	0.302
	$P = 0.92$	$P = 0.65$	$P = 0.39$	$P = 0.79$	$P = 0.21$
Cross-correlation	0.318	0.084	−0.225	−0.132	−0.005
	$P = 0.19$	$P = 0.73$	$P = 0.35$	$P = 0.59$	$P = 0.98$
Repertoire size	0.232	0.208	−0.114	−0.276	0.119
	$P = 0.34$	$P = 0.39$	$P = 0.64$	$P = 0.25$	$P = 0.63$

$N = 19$ for all correlations

Table 2 Spearman rank correlations between cognitive measures

	Novel foraging	Color association	Color reversal	Spatial learning	Detour-reaching
(a) Correlations for male subjects only ($N = 19$ for all correlations)					
Novel foraging	–	0.041	−0.080	0.444	0.044
		$P = 0.87$	$P = 0.74$	$P = 0.06$	$P = 0.86$
Color association		–	**0.627**	0.207	−0.076
			$P = 0.004$	$P = 0.39$	$P = 0.76$
Color reversal			–	0.381	−0.410
				$P = 0.11$	$P = 0.08$
Spatial learning				–	−0.168
					$P = 0.49$
(b) Correlations for male and female subjects together (Sample sizes are given in parentheses)					
Novel foraging	–	0.071 (39)	−0.019 (41)	0.119 (40)	−0.162 (40)
		$P = 0.67$	$P = 0.91$	$P = 0.47$	$P = 0.32$
Color association		–	**0.533** (39)	−0.039 (38)	0.001 (38)
			$P < 0.001$	$P = 0.82$	$P = 0.99$
Color reversal			–	−0.032 (40)	0.002 (40)
				$P = 0.85$	$P = 0.99$
Spatial learning				–	−0.118 (39)
					$P = 0.48$

Bold lettering indicates a significant correlation ($P < 0.05$)

Principal component analyses

Principal component analysis of the male cognitive measures identified two principal components with eigenvalues greater than 1. PC1 had an eigenvalue of 2.028 and explained 40.6% of the overall variance, while PC2 had an eigenvalue of 1.142 and explained 22.8% of the variance. Table 3 shows the component loadings, which were positive for PC1 for all measures except detour-reaching. Color association and color reversal loaded particularly strongly. Table 4 shows correlations between the song measures and PC1 and PC2. None of the correlations are significant.

Principal component analysis of the male song data identified a single principal component with an eigenvalue greater than 1. This PC1 had an eigenvalue of 1.531 and explained 51.0% of the variance. The loadings were 0.825 for proportion copied, 0.871 for average cross-correlation, and 0.304 for repertoire size. Scores on the song PC1 were not significantly correlated with either cognition PC1 or

Table 3 Variable loadings for principal components of the cognition measures for the male tests

	PC1	PC2
Novel foraging	0.218	0.679
Color association	0.829	0.277
Color reversal	0.849	0.210
Spatial learning	0.598	−0.501
Detour-reaching	−0.462	0.555

Table 4 Spearman rank correlations between song measures and the two principal components combining the cognitive measures

	PC1	PC2
Percent matching	−0.140	0.169
	$P = 0.57$	$P = 0.49$
Average cross-correlation	−0.026	0.102
	$P = 0.92$	$P = 0.68$
Repertoire size	−0.048	0.400
	$P = 0.85$	$P = 0.09$

$N = 19$ for all correlations

cognition PC2, nor with any of the separate cognition measures ($P > 0.05$ in all cases).

Generalized linear mixed models

The GLMM analysis uncovered several associations between song and cognition that were not indicated by simple correlations (Table 5). All three of the song measures (proportion copied, average cross-correlation, and repertoire size) were significantly associated in the predicted direction with performance on color reversal and spatial learning, meaning that for these cognitive measures better-learned song predicted better cognitive performance. Conversely, all three song measures were significantly associated in the direction opposite to prediction with novel foraging and detour-reaching, meaning that for these cognitive measures better-learned song predicted worse cognitive performance. Finally, none of the song measures were significantly associated either positively or negatively with color association.

Table 5 Results of the generalized linear mixed model analyses

	Estimate	SE	*t*	*P*
(a) Proportion copied				
Novel foraging	0.308	0.116	2.667	0.008
Color association	0.150	0.112	1.34	0.182
Color reversal	−0.393	0.102	−3.867	0.0001
Spatial learning	−0.472	0.100	−4.711	<0.0001
Detour-reaching	0.534	0.121	4.422	<0.0001
Log-likelihood	−379.88			
(b) Average cross-correlation				
Novel foraging	0.252	0.109	2.311	0.021
Color association	0.075	0.105	0.707	0.479
Color reversal	−0.433	0.097	−4.486	<0.0001
Spatial learning	−0.508	0.095	−5.324	<0.0001
Detour-reaching	0.373	0.111	3.353	0.0008
Log-likelihood	−383.73			
(c) Repertoire size				
Novel foraging	0.046	0.016	2.798	0.005
Color association	0.022	0.016	1.364	0.173
Color reversal	−0.054	0.014	−3.766	0.0002
Spatial learning	−0.076	0.014	−5.399	<0.0001
Detour-reaching	0.068	0.017	4.037	<0.0001
Log-likelihood	−377.37			

Each subtable gives a model in which the specified song variable is used to predict each of the five cognitive measures

Discussion

The pairwise correlations between song measures and cognitive measures gave little evidence of any association between song and cognitive ability: Of the 15 pairwise correlations linking song measures with cognitive measures, none was significant. Correlations of the song measures with the cognitive principal components also gave little evidence of any association between song and cognition. PC1, which was mainly indicative of success in color association and color reversal, was not significantly correlated with any of the three song measures. Similarly, PC2, which indicated success in novel foraging and detour-reaching, also was not significantly correlated with any of the song measures. Song PC1, combining the three song measures, was not correlated with cognition PC1 or PC2. The correlation analyses thus all supported a null hypothesis of no association between song and cognition.

By contrast, the generalized linear mixed model analysis uncovered a complex pattern of associations between song and cognition. Presumably, the greater success of the GLMMs in identifying associations can be ascribed in large part to the control exercised in this analysis of nest of origin and neophobia effects. Even though this analysis produced evidence of a number of significant associations, the results still do not cleanly support the basic hypothesis that song learning ability signals cognitive ability. All our song measures are structured so that a higher score indicates better song learning (more notes copied, more accurate copying, larger repertoire size). The five cognitive measures are all structured so that a larger score indicates worse cognitive performance (more trials to reach success). Thus, the hypothesis that song signals cognitive ability predicts negative associations of song measures with cognitive measures. In the GLMM results (Table 5), one of our cognitive measures (color association) consistently showed no association with song, two (color reversal and spatial learning) consistently showed negative associations with song (as predicted), and two (novel foraging and detour-reaching) consistently showed positive associations with song (opposite to prediction). In summary, as many results counter the hypothesis as support it.

It is also instructive to compare the results of our GLMM analyses with the results of the two previous studies of song and cognition in song sparrows. Both the earlier studies used repertoire size of free-living adult males as the sole measure of song quality, and both measured cognitive ability as we did on scales in which higher scores indicate worse cognitive performance. The strongest result in Boogert et al. (2011a) was a negative correlation between repertoire size and detour-reaching score, directly opposite to what we found for these two variables in this study. A somewhat weaker result in Boogert et al. (2011a) was a positive correlation between repertoire size and color reversal score, again directly opposite to what we found. The chief result in Sewall et al. (2013) was a positive correlation between repertoire size and spatial learning score, once again directly opposite to our results. Thus, the significant results from our study that can be compared against past studies are not consistent with the results of those studies.

We incorporated hand-rearing and tutoring with recorded song in our study in order to allow direct measurement of success in song learning through comparison of the songs produced by our subjects to the model songs they were tutored with. The study produced two measures of song learning, proportion of notes matching model songs and the cross-correlation of copied notes with model notes, that are superior as measures of song learning to any song measures produced in previous studies of song and cognition. At the same time, our study produced a measure of song complexity, repertoire size, that is identical to that used in some of the earlier studies of song and cognition and analogous to the others (Boogert et al. 2008, 2011a; Sewall et al. 2013; Farrell et al. 2012). Hand-rearing also allowed substantial control over our subjects' past experience, which Rowe and Healy (2014) warn is a potential source of bias in cognitive tests, particularly in tests of color association. All of our subjects were within a few days of each other in age, had spent their entire lifetimes in a protected laboratory environment, and had limited exposure to colorful objects. Variation in individual experience must still have existed but would be minimal compared to that in wild-caught birds.

One drawback of our study design is that the conditions under which our subjects learned songs were in some respects unnatural. Our subjects were not exposed to song models after 3 months post-hatching, and late exposure to song is known to affect song learning at least in western song sparrows (Nordby et al. 2001; Nulty et al. 2010). Our subjects also lacked the opportunity to interact with adult singers, and such interaction is known to affect aspects of song learning such as choice of models (Beecher and Burt 2004; Beecher et al. 2007). Perhaps as a consequence, the mean repertoire size of the males in our study (6.4 song types) was considerably lower than estimates of repertoire size for free-living birds in the source population (e.g., 7.9 in Hughes et al. 1998; 8.4 in Peters et al. 2000). Social isolation can also impair performance on cognitive tasks such as reversal learning (Schrijver et al. 2004), but our hand-reared subjects actually performed similarly in most cognitive tests to song sparrows captured from the wild as adults (Boogert et al. 2011a). A second drawback is that the hand-reared subjects were not exposed to many of the stresses that would be expected to occur in nature, such as food shortages, parasitism, and disease, and which might affect song development and cognitive development in parallel ways (Nowicki et al. 2002a; Pravosudov et al. 2005; Spencer and MacDougall-Shackleton 2011). A third drawback is that the demands of hand-rearing limited the sample size of males available for cognitive testing in our study. Consequently, our sample ($N = 19$) was smaller than the sample of wild-caught adult song sparrows that Boogert et al. (2011a) used in their novel foraging, color association, and color reversal tests ($N = 51$), but none of those tests gave results that supported the hypothesis that song signals cognitive ability. Our sample size was similar to that used by Boogert et al. (2011a) in their detour-reaching task ($N = 22$), which was the sole test that gave a positive result in that study. The remaining song sparrow study, by Sewall et al. (2013), used a smaller sample of wild-caught adults for a spatial learning task ($N = 14$), and found a significant association countering the hypothesis.

Rowe and Healy (2014) also warn of the difficulty of formulating behavioral tests that measure abilities that are unambiguously cognitive. Cognitive abilities are those having to do with information processing (Rowe and Healy 2014), and encompass in particular learning, memory, and decision-making (Shettleworth 2010). The cognitive tests administered in this study were explicitly designed to meet these criteria. The color association task is a straightforward test of associative learning, the ability to form a mental connection between two or more stimuli (Shettleworth 2010). The color reversal task is also an associative learning task, but in addition tests for the kind of behavioral flexibility that is important to innovation (Timmermans et al. 2000; Bond et al. 2007). Spatial learning is clearly a type of learning, but is different enough from other types of learning to be promoted as representing a separate cognitive module (Gallistel 2000). Detour-reaching is a kind of inhibitory-control task, in that subjects learn to inhibit a default response that is not effective under the conditions of the task. Performance on inhibitory-control tasks is typically associated with problem-solving ability (Diamond 1990; Dempster 1992; Carlson and Moses 2001) and is correlated with absolute brain volume across birds and mammals (MacLean et al. 2014). Of the five tasks, only novel foraging has an element of procedural knowledge, or "knowing how," which can be considered to be non-cognitive (Shettleworth 2010). This task does, however, require subjects to continue seeking rewards that are covered by lids so that they can no longer be sensed directly, a step that may be cognitively challenging.

If both the study design and the cognitive measures we used were valid, why then did we fail to find convincing evidence that song and cognition are positively associated overall? One obvious problem is with our initial assumption that cognitive measures in general would be positively associated with each other. Correlations between performance scores for different tasks within each round of male tests tended to be low; only 6 of 10 values were positive, and the overall mean was only 0.10. The one correlation that was nominally significant was between color association and color reversal; these were also the only two measures found to be significantly associated in an earlier study of the same species using four of the same five tasks (Boogert et al. 2011a). As these are the two cognitive tasks that are most similar in procedures and goals, it is perhaps not surprising that they show the strongest correlations. Analysis of the combined male and female data set confirmed with a larger sample size that correlations between the different cognitive measures are low. Only five of ten correlations were positive, and the mean correlation was only 0.04.

The general result that different cognitive measures are not consistently correlated with each other seems to be typical of bird studies (e.g., Keagy et al. 2011; Boogert et al. 2011a; Nettle et al. 2015; Shaw et al. 2015; Farrell et al. 2016) though spotted bowerbirds (*Chlamydera maculata*) provide an exception (Isden et al. 2013). Correlations between cognitive measures are also often low and inconsistent in sign in mammals other than humans (Locurto et al. 2003; Galsworthy et al. 2005), though again there are exceptions (Anderson 1993; Kolata et al. 2008). Only in humans are cognitive measures consistently correlated positively with one another in multiple studies (Carroll 1993). If, as in our study, different cognitive measures are not correlated with each other, then that removes one of the main reasons for expecting these measures to be correlated with song learning performance, while also making it unlikely that any one song measure will correlate strongly with multiple cognitive measures.

As detailed above, the present study does not replicate the conclusions of previous studies of song sparrows (Boogert et al. 2011a; Sewall et al. 2013) on positive and negative associations of specific cognitive abilities with song repertoire size. Similarly, one study of zebra finches found a positive association between a measure of song complexity and performance in a novel foraging task (Boogert et al. 2008), but the result was not confirmed by a second study of that species (Templeton et al. 2014). It is true that even studies that focus on the same species differ in some respects, for example in whether wild-caught (Boogert et al. 2011a, b; Sewall et al. 2013) or hand-reared subjects (this study) are used, or in whether subjects are tested solitarily (Boogert et al. 2008) or in groups (Templeton et al. 2014). The present study suggests (in the GLMM results) that three measures of song learning quality are positively associated with two specific cognitive abilities, color reversal learning and spatial learning. Further work will be needed to test whether these associations are replicable either within song sparrows or across songbirds as a whole.

Acknowledgements We thank Caitlin Cantrell, Philippa Tanford, and Nali Gillespie for help with hand-rearing and cognitive testing and Santiago Olivella and Casey Klofstad for assistance with the statistical analysis. We thank the National Science Foundation for supporting this work through Grants to SN (IOS-1144991) and WAS (IOS-1144995).

Compliance with ethical standards

Conflict of interest The authors declare they have no conflict of interest.

Ethical approval All procedures were approved by the Institutional Animal Care and Use Committee of Duke University (Protocol A032-14-02) and followed the Guidelines for the Treatment of Animals in Behavioural Research of the Animal Behavior Society and the Association for the Study of Animal Behaviour.

References

Anderson B (1993) Evidence from the rat for a general factor that underlies cognitive performance and that relates to brain size: intelligence? Neurosci Lett 153:98–102

Arcese P, Stoddard PK, Hiebert SM (1988) The form and function of song in female song sparrows. Condor 90:44–50

Bashi J (1977) Effects of inbreeding on cognitive performance. Nature 266:440–442

Bates D, Mächler M, Bolker BM, Walker SC (2015) Fitting linear mixed-effects models using lme4. J Stat Softw 67:1–48

Beecher MD, Brenowitz EA (2005) Functional aspects of song learning in songbirds. Trends Ecol Evol 20:143–149

Beecher MD, Burt JM (2004) The role of social interaction in bird song learning. Curr Dir Psychol Sci 13:224–228

Beecher MD, Campbell SE, Stoddard PK (1994) Correlation of song learning and territory establishment strategies in the song sparrow. Proc Nat Acad Sci USA 91:1450–1454

Beecher MD, Burt JM, O'Loghlen AL, Templeton CN, Campbell SE (2007) Bird song learning in an eavesdropping context. Anim Behav 73:929–935

Bond AB, Kamil AC, Balda RP (2007) Serial reversal learning and the evolution of behavioral flexibility in three species of North American corvids (*Gymnorhinus cyanocephalus, Nucifraga columbiana, Aphelocoma californica*). J Comp Psychol 121:372–379

Boogert NJ, Giraldeau LA, Lefebvre L (2008) Song complexity correlates with learning ability in zebra finch males. Anim Behav 76:1735–1741

Boogert NJ, Anderson RC, Peters S, Searcy WA, Nowicki S (2011a) Song repertoire size in male song sparrows correlates with detour reaching, but not with other cognitive measures. Anim Behav 81:1209–1216

Boogert NJ, Fawcett TW, Lefebvre L (2011b) Mate choice for cognitive traits: a review of the evidence in nonhuman vertebrates. Behav Ecol 22:447–459

Carlson SM, Moses LJ (2001) Individual differences in inhibitory control and children's theory of mind. Child Dev 72:1032–1053

Carroll JB (1993) Human cognitive abilities. Cambridge University Press, Cambridge

Dempster FN (1992) The rise and fall of the inhibitory mechanism: toward a unified theory of cognitive development and aging. Dev Rev 12:45–75

Diamond A (1990) Developmental time course in human infants and infant monkeys, and the neural bases of, inhibitory control in reaching. Ann NY Acad Sci 608:637–669

Eens M, Pinxten R, Verheyen RF (1991) Male song as a cue for mate choice in the European starling. Behaviour 116:210–238

Farrell TM, Weaver K, An YS, MacDougall-Shackleton SA (2012) Song bout length is indicative of spatial learning in European starlings. Behav Ecol 23:101–111

Farrell TM, Morgan A, MacDougall-Shackleton SA (2016) Developmental stress impairs performance on an association task in male and female songbirds, but impairs auditory learning in females only. Anim Cogn 19:1–14

Gallistel CR (2000) The replacement of general-purpose learning models with adaptively specialized learning modules. In: Gazzaniga MS (ed) The cognitive neurosciences. MIT Press, Cambridge, pp 1179–1191

Galsworthy MJ, Paya-Cano JL, Liu L, Monleón S, Gregoryan G, Fernandes C, Schalkwyk LC et al (2005) Assessing reliability, heritability and general cognitive ability in a battery of cognitive tasks for laboratory mice. Behav Gen 35:675–692

Hughes M, Nowicki S, Searcy WA, Peters S (1998) Song-type sharing in song sparrows: implications for repertoire function and song learning. Behav Ecol Sociobiol 42:437–446

Isden J, Panayi C, Dingle C, Madden J (2013) Performance in cognitive and problem-solving tasks in male spotted bowerbirds does not correlate with mating success. Anim Behav 86:829–838

Jacobs LF (1996) Sexual selection and the brain. Trends Ecol Evol 11:82–86

Keagy J, Savard JF, Borgia G (2009) Male satin bowerbird problem-solving ability predicts mating success. Anim Behav 78:809–817

Keagy J, Savard JF, Borgia G (2011) Complex relationship between multiple measures of cognitive ability and male mating success in satin bowerbirds, *Ptilonorhynchus violaceus*. Anim Behav 81:1063–1070

Kolata S, Light K, Matzel LD (2008) Domain-specific and domain-general learning factors are expressed in genetically heterogeneous CD-1 mice. Intelligence 36:619–629

Li NP, Bailey JM, Kenrick DT, Linsenmeier JAW (2002) The necessities and luxuries of mate preferences: testing the trade-offs. J Pers Soc Psychol 82:947–955

Locurto C, Fortin E, Sullivan R (2003) The structure of individual differences in heterogeneous stock mice across problem types and motivational systems. Genes Brain Behav 2:40–55

MacLean EL, Hare B, Nunn CL, Addessi E, Amici F, Anderson RC, Aureli F et al (2014) The evolution of self-control. Proc Nat Acad Sci USA 111:E2140–E2148

Marler P, Mundinger P (1971) Vocal learning in birds. In: Moltz H (ed) Ontogeny of vertebrate behavior. Academic Press, New York, pp 389–450

Marler P, Peters S (1987) A sensitive period for song acquisition in the song sparrow, *Melsopiza melodia*: a case of age-limited learning. Ethology 76:89–100

Marler P, Peters S (1988) The role of song phonology and syntax in vocal learning preferences in the song sparrow, *Melospiza melodia*. Ethology 77:125–149

Matzel LD, Han YR, Grossman H, Karnik MS, Patel DN, Scott N, Specht SM et al (2003) Individual differences in the expression of a "general" learning ability in mice. J Neurosci 23:6423–6433

Miller G (2000) The mating mind: How sexual choice shaped the evolution of human nature. Doubleday, New York

Nettle D, Andrews CP, Monaghan P, Brilot BO, Bedford T, Gillespie R, Bateson M (2015) Developmental and familial predictors of adult cognitive traits in the European starling. Anim Behav 107:239–248

Nordby JC, Campbell SE, Beecher MD (1999) Ecological correlates of song learning in song sparrows. Behav Ecol 10:287–297

Nordby JC, Campbell SE, Beecher MD (2001) Late song learning in song sparrows. Anim Behav 61:835–846

Nowicki S, Hasselquist D, Bensch S, Peters S (2000) Nestling growth and song repertoire size in great reed warblers: evidence for song learning as an indicator mechanism in mate choice. Proc Biol Sci 267:2419–2424

Nowicki S, Searcy WA, Peters S (2002a) Brain development, song learning and mate choice in birds: a review and experimental test of the "nutritional stress hypothesis". J Comp Phys A 188:1003–1014

Nowicki S, Searcy WA, Peters S (2002b) Quality of song learning affects female response to male bird song. Proc Biol Sci 269:1949–1954

Nulty B, Burt JM, Akçay Ç, Templeton CN, Campbell SE, Beecher MD (2010) Song learning in song sparrows: relative importance of autumn vs. spring tutoring. Ethology 116:653–661

O'Rourke N, Hatcher L (2013) Factor analysis and structural equation modeling. SAS Institute, Cary

Peters S, Searcy WA, Beecher MD, Nowicki S (2000) Geographic variation in the organization of song sparrow repertoires. Auk 117:936–942

Peters S, Searcy WA, Nowicki S (2014) Developmental stress, song learning, and cognition. Integr Comp Biol 54:555–567

Plomin R (2001) The genetics of G in humans and mouse. Nat Rev Neurosci 2:136–141

Pravosudov VV, Lavenex P, Omanska A (2005) Nutritional deficits during early development affect hippocampal structure and spatial memory later in life. Behav Neurosci 119:1368–1374

Prokosch MD, Coss RG, Scheib JE, Blozis SA (2009) Intelligence and mate choice: intelligent men are always appealing. Evol Hum Behav 30:11–20

R Core Team (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Reid JM, Arcese P, Cassidy ALEV, Hiebert SM, Smith JNM, Stoddard PK, Marr AB et al (2004) Song repertoire size predicts initial mating success in male song sparrows, *Melospiza melodia*. Anim Behav 68:1055–1063

Reid JM, Arcese P, Cassidy ALEV, Marr AB, Smith JNM, Keller LF (2005) Hamilton and Zuk meet heterozygosity? Song repertoire size indicates inbreeding and immunity in song sparrows (*Melospiza melodia*). Proc Biol Sci 272:481–487

Riebel K (2009) Song and female mate choice in zebra finches: a review. Adv Study Behav 40:197–238

Rowe C, Healy SD (2014) Measuring variation in cognition. Behav Ecol 25:1287–1292

Schrijver NCA, Pallier PN, Brown VJ, Würbel H (2004) Double dissociation of social and environmental stimulation on spatial learning and reversal learning in rats. Behav Brain Res 152:307–314

Searcy WA (1984) Song repertoire size and female preferences in song sparrows. Behav Ecol Sociobiol 14:281–286

Searcy WA, McArthur PD, Yasukawa K (1985) Song repertoire size and male quality in song sparrows. Condor 87:222–228

Sewall KB, Soha JA, Peters S, Nowicki S (2013) Potential trade-off between vocal ornamentation and spatial ability in a songbird. Biol Lett 9:20130344

Shaw RC, Boogert NJ, Clayton NS, Burns KC (2015) Wild psychometrics: evidence for 'general' cognitive performance in wild New Zealand robins, *Petroica longipes*. Anim Behav 109:101–111

Shettleworth SJ (1998) Cognition, evolution, and behavior. Oxford University Press, Oxford

Shettleworth SJ (2010) Cognition, evolution, and behavior, 2nd edn. Oxford University Press, Oxford

Shohet AJ, Watt PJ (2009) Female guppies *Poecilia reticulata* prefer males that can learn fast. J Fish Biol 75:1323–1330

Spearman C (1904) "General intelligence" objectively determined and measured. Am J Psychol 15:201–292

Spencer KA, MacDougall-Shackleton SA (2011) Indicators of development as sexually selected traits: the developmental stress hypothesis in context. Behav Ecol 22:1–9

Spritzer MD, Meikle DB, Solomon NG (2005) Female choice based on male spatial ability and aggressiveness among meadow voles. Anim Behav 69:1121–1130

Templeton CN, Laland KN, Boogert NJ (2014) Does song complexity correlate with problem-solving performance in flocks of zebra finches? Anim Behav 92:63–71

Timmermans S, Lefebvre L, Boire D, Basu P (2000) Relative size of the hyperstriatum ventrale is the best predictor of feeding innovation rate in birds. Brain Behav Evol 56:196–203

Vyas A, Harding C, Borg L, Bogdan D (2009) Acoustic characteristics, early experience, and endocrine status interact to modulate female zebra finches' behavioral responses to songs. Horm Behav 55:50–59