Female zebra finches prefer the songs of males who quickly solve a novel foraging task to the songs of males unable to solve the task

Clara Howell[1], Rindy Anderson[2], and Elizabeth Derryberry[3]

[1]University of Tennessee
[2]Florida Atlantic University - Davie Campus
[3]Tulane University

May 5, 2020

Abstract

Correlative evidence suggests that high problem-solving and foraging abilities in a mate are associated with direct fitness advantages, so it would benefit females to prefer problem-solving males. Recent work has also shown that females of several bird species who directly observe males prefer those that can solve a novel foraging task over those that cannot. In addition to or instead of direct observation of cognitive skills, many species utilize assessment signals when choosing a mate. Here we test whether females can select a problem-solving male over a non-solving male when presented only with a signal known to be used in mate assessment: song. Using an operant conditioning assay, we compared female zebra finch (Taeniopygia guttata) preference for the songs of males that could quickly solve a novel foraging task to the songs of males that could not solve the task. Females were never housed with the test subject males whose song they heard, and the only information provided about the males was their song. We found that females elicited more songs of problem-solving males than of non-solvers, indicating that song can contain information about a male's ability to solve a novel foraging task and that naïve females prefer the songs of problem-solving males.

Introduction :

A key question in the study of animal cognition is whether sexual selection has contributed to its evolution (Darwin 1871; Boogert, Fawcett, et al. 2011). A related but subtly different question is whether signals used in mate choice reflect cognitive abilities that are under natural selection, and would therefore offer direct or indirect benefits in a potential mate. The extension of this question is whether the utility of certain assessment signals in selecting a mate with beneficial cognitive abilities has led to their continued use and evolution.

The hypothesis that signals used in mate choice ("assessment signals") serve as indicators of male quality has been extensively supported in a wide variety of taxa (reviewed in Wilgers and Hebets 2015), including fish (Houde and Torio 1992; Karino et al. 2005), birds (Hill 1991; Doucet and Montgomerie 2003), insects (David et al. 2000; Holzer et al. 2003), and even humans (Hume and Montgomerie 2001; Little et al. 2008). The hypothesis that assessment signals can serve as an indicator of general brain function has also been supported (Boogert, Fawcett, et al. 2011), in particular by studying the effects of developmental stress on bird song (Peters et al. 2014). Developmental stress lowers cognitive performance in a range of taxa (Levitsky and Strupp 1995; Erhard et al. 2004; Kitaysky et al. 2006; Santos de Souza et al. 2008; Peters et al. 2014; Kriengwatana et al. 2015) and, specifically within birds, on a range of tasks including novel foraging tasks (Kitaysky et al. 2006), spatial learning tasks (Pravosudov et al. 2005; Kriengwatana et al. 2015), associative learning tasks (Fisher et al. 2006; Farrell et al. 2016), and auditory learning tasks (Farrell et al. 2016).

1

Developmental stress also negatively impacts preferred features of song, such as complexity (Spencer et al. 2003; Buchanan et al. 2004), bout duration (Buchanan et al. 2003), and accuracy in relation to tutor song (Nowicki et al. 2002), and is less preferred by females when presented in comparison with song from developmentally robust males (Spencer et al. 2005; Searcy et al. 2010).

These findings suggest that song can serve as an indicator of gross brain quality and function in the case of developmental inequity between potential mates. However, in the absence of conditions like developmental stress or genetic abnormalities that cause widespread disruption in the body and brain, there is increasing evidence that avian intelligence is more modular than general in nature (Searcy and Nowicki 2019). For instance, there are seldom correlations between performance on different cognitive tasks within individuals (Nettle et al. 2015; van Horik and Madden 2016; Anderson et al. 2017; Medina-García et al. 2017; DuBois et al. 2018; MacKinlay and Shaw 2018), with some exceptions (Shaw et al. 2015; Ashton et al. 2018). Notably, exceptions tend to occur in wild populations where unequal developmental conditions are more likely. Because of the apparent modularity of avian intelligence, it is unclear whether song is a reliable indicator of the specific cognitive skills that would offer females fitness benefits if found in a potential mate.

One such cognitive skill is foraging ability. Foraging efficiency has been shown to increase fitness in a range of avian species (Orians 1969; Weathers and Sullivan 1989; Lemon and Barth 1992; Lescroël et al. 2010; Cole et al. 2012), and is a particularly important skill in a mate in monogamous species that engage in biparental food provisioning. Indicative of its importance, foraging ability has also been shown to directly affect mate preferences in crossbills (*Loxia curvirostra*), budgerigars (*Melopsittacus undulatus*), and zebra finches (*Taeniopygia guttata*) when females directly observe males solving a foraging task (Chantal et al. 2016; Chen et al. 2019) or more efficiently extracting food from a source (Snowberg and Benkman 2009; Chantal et al. 2016). Foraging ability is also correlated with problem-solving on artificial tasks in the wild (Cole et al. 2012; Cauchard et al. 2013), which has in turn been correlated with reproductive success (Keagy et al. 2009; Cole et al. 2012; Cauchard et al. 2013; Ashton et al. 2018).

The ability to solve a novel foraging problem is thus related to reproductive success and has been shown in some species to be preferred by females when observed directly. However, it remains unclear whether the same signals that can indicate gross brain function, as determined by developmental conditions, can reflect this specific skill. Do songs contain information about male foraging ability, given the apparent modularity of avian cognition, and could this help explain the utility of song as an assessment signal? Or, conversely, does song only contain information about gross developmental conditions, and not more specific skills that would offer females direct fitness benefits?

To answer these questions, we tested whether female zebra finches would discriminate between males with superior versus inferior performance on a novel foraging task based solely on their song. Using an operant conditioning assay, we compared female preference for the songs of males that quickly solved the novel foraging task to the songs of males that were incapable of solving the task. The females had no exposure to the males whose songs they were tested with, thus isolating their knowledge of the males to information contained in their songs. We predicted that if song is a reliable signal of novel foraging skill, females would be able to identify and prefer songs from males who were capable of quickly solving a novel foraging task over songs from males who were unable to solve the task.

Methods:

Subjects and Housing Conditions:

To assay male foraging ability and acquire recordings of male songs, we obtained 25 adult male zebra finches of unspecified age from Magnolia Farms avian breeder (Anaheim, CA). Birds were housed in a group cage for four days upon arrival, then weighed with a Pesola scale and placed in individual cages in a group room. Cages were wire and measured 48cm x 25cm x 30cm with two perches and one cuttlebone each. Food and water were provided ad libitum. Vivarium rooms were illuminated on a 13:11 light-dark cycle. Temperature in the room was maintained at 22°C. Housing conditions were approved by Tulane IACUC Protocol 0427R.

To measure female preference, we obtained 15 adult female zebra finches of unspecified age, also from Magnolia Farms. Females were purchased 18 months after the male zebra finches were purchased to minimize the possibility of prior interaction between males and females due to overlap at Magnolia Farms. Housing conditions were identical to the first group of birds. Males were not housed in the facility at the same time as the females, and females were not exposed to their songs prior to testing.

Novel foraging assay :

To solve the novel foraging task, males learned to remove lids covering baited wells in a block. The block was 10 x 14 cm and composed of grey composite plastic with six drilled wells (1.7 cm diameter x 1 cm depth). The food reward consisted of 2-3 millet seeds placed in the bottom of 4 of 6 wells. Baiting patterns were randomly generated with R (R Development Core Team 2016) and changed between each trial. Lids were made of blue and yellow round plastic counting chips glued to a round rubber bottom, such that lids fitted into the wells and needed to be lifted off in order for the bird to obtain the food reward. Stages consisted of: (1) just the block, (2) lids placed adjacent to, but not covering baited wells, (3) lids half-covering wells, (4) lids tipped into wells, and (5) lids completely covering and fitted into wells. Testing was done in the housing room and birds remained in auditory contact with their flockmates to alleviate stress caused by moving or social isolation. Dividers were placed between cages during testing so that neighbors of the study subject could not see the task beforehand.

Trials consisted of two-minute periods in which the block was placed in the cage and the bird was allowed to interact with it. After each two-minute trial the block was removed from the cage for approximately ten minutes while other birds were tested. Each bird was given six trials per day. Food was removed five hours before testing to increase motivation. After each round of testing, a motivation check was done in which a food dish was placed in the cage and the time it took the bird to approach and eat from it was recorded. If birds ate from food dishes in less than one minute, they were considered sufficiently motivated. All birds passed all motivation tests, indicating that food removal left them sufficiently motivated to obtain a food reward.

Birds had to eat from at least two baited wells to pass a trial, and had to pass three out of four consecutive trials to move on to the next stage. Passing criteria remained the same throughout the novel foraging task. If birds were stuck on a particular stage for more than four days, they were moved back to a previous stage or, for stage 1 only, the block was left in the cage overnight. If a bird did not complete a step (not the entire task, just one step) within 60 trials, it was removed from trials and designated a Non-Solver, as further food removal and testing was potentially harmful to birds and in our judgment these individuals were not likely to progress further.

Stimulus Sets and Song Recording:

The 6 best-performing males (those that learned the task in the fewest number of trials) were designated Solvers, while the 6 worst-performing males (those that did not solve the task) were designated Non-Solvers. 6 Stimulus sets were created by randomly pairing 1 Solver song with 1 Non-Solver song. All Solvers used as stimulus males solved the task in under 30 trials (mean=22.67 trials); all Non-Solvers took at least 60 trials before they were pulled from testing (mean=90.67 trials). The 6 fastest Solvers were chosen to maximize disparity between the abilities of Solvers and Non-Solvers. Five out of the six Non-Solvers dropped out during stage 5, in which they could no longer see the food reward. The remaining Non-Solver dropped out during stage 1, in which food was placed in wells. Because of concern that this Non-Solver was categorically different in ability than the 5 Non-Solvers who failed at the same stage, statistics were run on female preference including and not including this stimulus set. Stimulus Solver and Non-Solver males were recorded in sound attenuation chambers (Industrial Acoustics) using Shure SM57 directional microphones and Sound Analysis Pro (see Tchernichovski et al. 2000). Males were placed in divided cages with a female (not a study subject and not housed with study subjects) in the other half of the cage in order to elicit directed song (ten Cate 1985). Males and females were thus in visual and auditory contact but not physical contact. To acclimate, pairs were placed in sound attenuation chambers for 24 hours prior to recording. Song output was recorded

during the subsequent 24 hours. The majority of males produced hundreds of directed songs in this time period. Males that produced fewer songs were re-recorded in separate sessions with different females until enough songs for analysis were produced. Female zebra finches have been shown to prefer longer songs (Neubauer 1999), so we selected the longest song from each male's repertoire.

Zebra finch recordings were analyzed after the experiment using Raven Pro Interactive Sound Analysis Software (The Cornell Lab of Ornithology, Ithaca, NY, version 1.5; Bioacoustics Research Program 2014). Typically, song complexity is analyzed by selecting random motifs from across a male's repertoire and averaging their complexity values (e.g. Boogert et al. 2008). However we were not interested in whether problem-solving ability correlated with general song complexity in males, but in whether song complexity explained the female preference we observed. We thus analyzed only the songs used as stimuli. We measured several potential measures of complexity in this species: song length, number of phrases, average phrase length, average elements per phrase, and total unique elements per song. Data were collected by visual inspection of spectrograms (256 pt. transform, frequency resolution = 86.1 Hz) and elements were categorized following Airey and DeVoogd (2000). To categorize elements as same or different, we used characteristics of the number and distribution of harmonics, frequency modulation, and element duration. Introductory elements were counted as phrases but excluded from phrase length and elements per phrase analyses.

Female Choice Assay:

Prior to the mate choice assay, each female's cage was moved into a sound attenuation chamber equipped with one audio speaker. Females were given 24 hours to habituate to the chambers, at which point operant perches were installed on either side of the front cage door, so that total perches in the cage included two operant perches and two normal perches placed diagonally in the back of the cage. Operant perches were placed approximately 10 cm above the cage floor and 25 cm apart. Perches were approximately 9 cm long, $\frac{1}{2}$ cm in diameter, and made of wood. Perches were attached to Honeywell lever arm depression micro-switches. The perch on the left when facing the cage was designated perch A, and the perch on the right was designated perch B. Hopping on a perch caused the perch to lower slightly and trigger a pressure-sensitive button on the micro-switch, which in turn triggered playback of a song through the speaker. In this type of operant conditioning, the song itself serves as the reward and encourages females to continue eliciting song through hops (Stevenson 1967; Riebel 2000; Anderson 2009). All songs were volume-adjusted to 65 dB SPL at the chamber center. All operant data were collected through Sound Analysis Pro (see Tchernichovski et al. 2000).

Females learned to trigger song during an initial training period (Anderson et al. 2014). For training, females were presented with conspecific/ heterospecific song pairs (heterospecific song was from rufous-collared sparrows, *Zonotrichia capensis*). Whether perch A played conspecific and perch B played heterospecific, or vice versa, was alternated between birds, and switched every day for the same bird. Switching songs between perches helped control for side bias. In order to pass training, females had to hop on both perches for two days in a row. All females passed training within 4 days.

After training, females were presented with Solver/Non-Solver stimuli. Females were given a new pair of Solver and Non-Solver songs every day, in which a single song from a Solver male was triggered by a hop on one perch, and a single song from a Non-Solver male was triggered by a hop on another perch. The side that played Solver song was switched each day to control for side bias. Each bird heard 4 of 6 Solver/Non-Solver pairs, as female response tends to drop off after prolonged periods in preference testing. Testing multiple stimulus sets on each female also greatly reduces the risk that a chosen stimulus set is non-representative of its category and yet repeatedly measured (i.e. pseudoreplication in acoustic experiments; Kroodsma 1989). Trials ran from 11am to 5pm each day, during which the number of hops on each perch was recorded. Outside of trial times, hops on a perch did not trigger playback.

Statistical Analysis:

Preference data are typically analyzed using non-parametric analyses, such as the Wilcoxon sign-rank test, in which preferences are expressed as ratios and compared to a null. However this approach has several drawbacks. For example, (1) it does not measure the strength of preferences, only whether they are statistically

different from the null, (2) it fails to account for different counts per replicate and does not adjust confidence intervals accordingly, and (3) it does not model variation in individual-level choice—for instance, if half of individuals in a population strongly prefer one option and half another, the population preference will still be 0.5 and may not differ significantly from the null (Fordyce et al. 2011). Given these drawbacks, we chose to use a hierarchical Bayesian model designed for ecological preference and count data, implemented using the R packages bayespref (version 1.0) and coda (version 1.0) (Plummer et al. 2006; Fordyce et al. 2011). The benefit of the hierarchical Bayesian model is that it directly estimates the strength of preference, appropriately models uncertainty, and gives estimates of both individual and population-level preferences.

The R package bayespref uses MCMC chains to first assess the probability of the experimental count data given an individual preference, thus modeling individual preference, and at each step in the chain further assesses the probability of an individual preference given a population preference, thus modeling population preference. In a two-choice scenario like ours, individual preferences are modeled as binomial distributions and population preferences as beta distributions. The first model was run using bayespref for female preference for conspecific vs. heterospecific song, to confirm that females were showing the expected preference for conspecific song and that the operant trials were reflecting preference. The second model was run for Solver vs. Non-Solver song to determine whether females showed a population-level preference for Solver song. Models were run with 10,000 generation MCMC chains with 1,000-generation burn-ins. Mixing of search chains was verified via diagnostic plots of MCMC samples and checking the effective sample size.

In the first model we used data for each female, where the total number of conspecific hops over two days was modeled as one possible outcome and the total number of heterospecific hops over two days was modeled as the alternative outcome. In the second model we also had data for each female, where the total number of Solver hops over all test days was modeled as one possible outcome and the total number of Non-Solver hops over all test days was modeled as the alternative outcome. This controlled for side bias, which in some females was noticeable even in conspecific vs. heterospecific trials. To evaluate the impact of stimulus set on female preference, and to ensure that no one stimulus set was disproportionately affecting results, we ran a third hierarchical Bayesian model using stimulus song at the individual level and preference for all stimulus songs in a given category (Solver or Non-Solver) at the population level. In this model we had data for each stimulus set $(1 - 6)$, where the total number of female hops across all trials by all birds for the Solver song was modeled as one possible outcome and the total number of female hops across all trials for the Non-Solver song was modeled as the alternative outcome. In the same way that the first model calculated individual female preference and confirmed that no one female was driving population-level preference, this model confirmed that no one song set was disproportionately "attractive" and driving preference for Solver song.

Because of concern that one of the Non-Solvers was categorically different in ability than the other 5, we also ran these Bayesian models using only data from the 5 stimulus sets in which the Non-Solvers all failed at the same step.

Although we considered the Bayesian approach sufficient to answer our research questions, we also ran a mixed effects linear model using the preference data for each trial and including stimulus ID, female ID, and order of presentation as random effects. The results of this model can be found in Supplementary Materials.

We further tested whether any measures of song complexity (song length, number of phrases, phrase length, number of elements, or number of unique elements) differed between stimulus pairs. To test this we used paired t-tests. We corrected for multiple testing (here, five song variables) using a Bonferroni correction, such that the adjusted alpha was 0.01.

Because we were interested in whether neophobia affected performance, we also tested whether Non-Solvers took longer than Solvers to solve the first stage of the task in which they were introduced to the block, and the second stage in which they were introduced to the lids, using an independent t-test. All statistical analyses were performed in R (Version 3.5.0; R Development Core Team 2016).

Results

5

Novel foraging task results

25 males were assessed with the novel foraging assay, of which 7 were unable to complete the task within the established cut-off (birds that did not pass any given stage within 60 trials were pulled from testing). Of those males that did pass, it took an average of 43.61 trials (s.d.=22.40 trials) to do so. Solvers used for stimuli were selected from the 6 fastest males to solve the task, and took an average of 22.67 trials to solve the task. Six out of seven non-solvers were used for stimuli, and all Non-Solvers took at least 60 trials before they were pulled from testing (average=90.67 trials). Full results from the novel foraging task can be found in the Supplementary Materials.

Females preferred conspecific song to heterospecific song

The first Bayesian model (*Fig. 1*) confirmed that females show a preference for conspecific song in the operant chambers. Population preference for conspecific song was 0.705 (95% credible intervals: 0.687, 0.748) and population preference for heterospecific song was 0.295 (95% credible intervals: 0.252, 0.313). Effective sample size (ESS), a measure of the MCMC chain function and not associated with sample size in a typical non-parametric analysis, was 1593.247. Average hops per conspecific/heterospecific trial was 521.93 (s.d. = 300.96). Full results from the conspecific/heterospecific preference trials can be found in the Supplementary Materials.

Females preferred songs produced by males that solved the novel foraging task

The second Bayesian model (*Fig. 2*) showed that females prefered Solver song and gave information about strength of preference. Population preference for Solver song was 0.594 (95% credible intervals: 0.581, 0.628), and population preference for Non-Solver song was 0.406 (95% credible interval: 0.372, 0.419). ESS was 1674.475. In a Bayesian preference model, confidence that one group is preferred over another increases as overlap in the confidence intervals of both groups declines. In our model there was no overlap between the 95% credible intervals for Solver preference and Non-Solver preference, indicating a substantially higher preference for the song of Solver males, with a 5% or lower probability that Non-Solver song was preferred equally or more than Solver song (Fordyce et al. 2011). For comparison, 95% credible intervals for group conspecific song preference were 0.687, 0.748; and 95% credible intervals for group heterospecific song preference were 0.252, 0.313 (see *Fig. 1*). There was a greater difference between conspecific and heterospecific confidence intervals, but preference for conspecific song was only approximately 11 percentage points higher (0.705) than preference for Solver song (0.594).

These patterns remained consistent when the 6th stimulus set was dropped (Solver preference: 0.581, 95% credible interval: 0.567, 0.617; Non-Solver preference: 0.419, 95% credible interval: 0.383, 0.433).

In Solver/Non-Solver trials, the average number of hops per trial was 704.84 (s.d. = 384.59). Full results from Solver/Non-Solver trials can be found in the Supplementary Materials.

Females preferred the Solver song in all stimulus sets

The third Bayesian model (*Fig. 3*) confirmed a general preference for Solver songs and showed that in each stimulus set, Solver song was preferred over Non-Solver song. Aggregate preference for Solver song was 0.608 (95% confidence intervals: 0.592, 0.625). Aggregate preference for Non-Solver song was 0.392 (95% confidence intervals: 0.374, 0.408). In the stimulus set that was furthest apart in preference, Solver song had an estimated preference of 0.682 (95% confidence intervals: 0.669, 0.695) while Non-Solver song had an estimated preference of 0.318 (95% confidence intervals: 0.305, 0.331). In the stimulus set that was closest in preference, Solver song had an estimated preference of 0.523 (95% confidence intervals: 0.551, 0.535), while Non-Solver song had an estimated preference of 0.477 (95% confidence intervals: 0.564, 0.489). Even in the closest stimulus set, 95% confidence intervals did not overlap (thus making it a 5% or less chance that the Non-Solver song was preferred over the Solver song). We can therefore say that Solver song was preferred in all stimulus pairs and that population preference for Solver song (see *Fig. 2*) was not driven by one or two particularly attractive Solver songs, but by a consistent pattern in preference across all six stimulus sets.

6

These patterns remained consistent when the 6thstimulus set was dropped (Solver preference by stimulus set: 0.597, 95% credible interval: 0.583, 0.613; Non-Solver preference by stimulus set: 0.403, 95% credible interval: 0.387, 0.417).

Song complexity does not appear to explain female preference for Solver song

Solver and Non-Solver songs did not differ significantly in any measures of complexity: song length (average Solver song length=4.5 seconds, average Non-Solver song length=3.97 s,; T=0.384, DF=5, P=0.461), number of phrases (average Solver phrases=5.83, average Non-Solver phrases=4,; T=3.84, DF=5, P=0.012), phrase length (average Solver phrase length=0.834 s, average Non-Solver phrase length=0.844 s; T= -0.052, DF=5, P=0.961), number of elements per phrase (average Solver elements=5.25, average Non-Solver elements=4.92,; T=0.353, DF=5, P=0.739), and number of unique elements (average Solver unique elements=5.33, average Non-Solver unique elements=5.5,; T=-0.143, DF=5, P=0.892). However, our power to determine an effect of one standard deviation given an alpha of 0.01 is 0.194, so we cannot say with confidence that Solvers and Non-Solvers in general do not have a difference in song complexity.

Neophobia does not appear to explain faster learning by Solvers

Five of six Non-Solvers ate consistently from the baited wells of a novel object (stage 1). The Non-Solver who did not eat from the grid was given a maximum score of 60 in comparisons. Non-Solvers did not significantly differ from Solvers in the number of trials required to pass stage 1 (T= -1.58, DF=5.1, P=0.173). However, Non-Solvers did tend to require more trials to pass stage 1 than Solvers (Solver mean=6 trials, Non-Solver mean=19.5 trials. Solvers did not differ from Non-Solvers in the number of trials required to pass stage 2, in which the second group of novel objects, colorful lids, was presented (T= -1.27, DF=4, P=0.273). Our power to determine an effect of 0.5 standard deviations given an alpha of 0.05 is 0.635, so we cannot fully exclude differences in neophobia between the two groups as a contributor to female preference.

Discussion

In our operant assay, females hopped on perches that triggered songs from males that completed a novel foraging task more frequently than they triggered songs from males unable to complete the task, indicating that the females preferred the songs of Solvers over the songs of Non-Solvers. Past work has shown a strong correlation between song preferences measured via operant conditioning and copulation solicitation displays (Anderson 2009), the latter being considered a fairly close proxy for mating preference (Holveck and Riebel 2007). Thus our findings support the hypothesis that females prefer the assessment signal of a male better at a novel foraging task, and replicate patterns of preference found when females directly observe males completing a foraging task (Snowberg and Benkman 2009; Chantal et al. 2016; Chen et al. 2019).

Our results suggest that if females chose to mate with the males whose song they preferred, they would be selecting Solver males more often than Non-Solver males. However, the mechanism linking problem-solving and song quality is still unknown. Although song was recorded after the novel foraging task, we think it highly unlikely that performing the task affected song, particularly as zebra finches have a single song that is crystallized in adolescence (Zann 1996). Rather, the more plausible interpretation of our data is that a third factor is correlated with both song quality and performance on the novel foraging task, such as a specific cognitive ability or personality trait. We are unable to speculate about this factor because the causes of successful learning on artificial problem-solving tasks remain murky and debated (Rowe and Healy 2014). Performance on artificial novel-foraging assays is thought to measure cognitive ability, but can also be affected by personality factors such as neophobia, boldness, and individual differences in motivation (reviewed in Griffin et al. 2015). We assured that all males were food-motivated by depriving them of food before testing and also confirmed that all males were sufficiently motivated to eat by measuring latency to feed once food was returned to the home cage, thus we do not think that differences in food motivation can explain our results. We also examined one measure of neophobia, latency to approach a novel object (Bouchard et al. 2007; Cauchard et al. 2013; Shaw et al. 2015), and found no statistically significant differences in neophobia between Solvers and Non-Solvers. Given the power of our analysis, we cannot definitively say that Solvers were less neophobic. However, this finding is consistent with a recent meta-analysis of personality and cognition

in birds that found no correlation between fear of novel objects and performance on novel foraging assays across multiple species (Dougherty and Guillette 2018). Thus while we cannot rule out the possibility that motivation or neophobia affected performance on the task and that females were responding to differences in personality that were reflected in song, we suggest that the more likely interpretation of our data is that the novel foraging task measured some aspect of cognitive ability that was also reflected in song.

While females were evidently responding to differences in Solver vs. Non-Solver song, how that information is encoded in song remains unknown. Past research has examined whether features of song, such as complexity (Boogert et al. 2008; Templeton et al. 2014), repertoire size (Boogert, Anderson, et al. 2011; Sewall et al. 2013; MacKinlay and Shaw 2018), and species typicality (DuBois et al. 2018) correlate with cognitive performance. The majority of these (Boogert, Anderson, et al. 2011; Templeton et al. 2014; DuBois et al. 2018; MacKinlay and Shaw 2018) found predominately null relationships between song macrofeatures and cognitive performance, and two (Boogert, Anderson, et al. 2011; Sewall et al. 2013) even found a significant inverse relationship. We measured the complexity of stimulus songs and found no statistically significant differences in song length, number of phrases, phrase length, number of elements, or number of unique elements between the stimulus songs produced by Solver and Non-Solver males. Given the power of our analysis we cannot say that problem-solving ability is unrelated to song complexity in zebra finches in general, only that there was no significant relationship within our sample of stimulus songs. The finding that females significantly preferred Solver songs despite no apparent difference in song complexity suggests that females attended to some feature in the songs that we as researchers could or did not measure, as has been found before in assays of local vs. foreign song (Anderson et al. 2014). Differences between the songs sung by the two groups of males could include variations in fine acoustic structure, such as minute changes in relative amplitude of harmonics and periods between amplitude peaks within an element, which have been shown to encode biologically relevant information in zebra finch calls (Prior et al. 2018). This study further highlights the need for future studies of female preference to measure female response to stimuli in addition to quantifying and categorizing stimuli attributes, as perhaps more subtle differences in tone or note structure encode information about male attributes.

Although these results are consistent with the hypothesis that female songbirds could select better foragers as mates by assessing features of song, future work in wild populations is needed to confirm that male performance on artificial novel foraging tasks is correlated with foraging efficiency. While performance on another artificial measure of problem-solving has been correlated with foraging efficiency in the wild (Cole et al. 2012), the specific lid-flipping task that we used to measure novel foraging ability has not been. And while these results indicate that song may be a reliable signal of a fitness-relevant cognitive task in the zebra finch, the songbird clade contains a huge variety of song types and singing behaviors (Catchpole and Slater 2008). It is worth investigating whether females of other species also prefer songs from males who are better foragers, or whether certain species are able to encode this type of information better than others.

Ethics: Ethical approval for this research was obtained from the Tulane University Animal Use Committee, Protocol 0427R.

Data availability: Available on Dryad.

Competing interests: We declare no competing interests.

Author contributions: C.H., R.A., and E.D. conceived of and designed experiment. C.H. and E.D. collected data and performed analyses. C.H. wrote the manuscript with support from R.A. and E.D.

References

Airey DC, DeVoogd TJ. 2000. Greater song complexity is associated with augmented song system anatomy in zebra finches. Neuroreport. 11:2339–2344. doi:10.1097/00001756-200007140-00054.

Anderson RC. 2009. Operant conditioning and copulation solicitation display assays reveal a stable preference for local song by female swamp sparrows Melospiza georgiana. Behav Ecol Sociobiol. 64:215–223. doi:10.1007/s00265-009-0838-y.

Anderson RC, Peters S, Nowicki S. 2014. Effects of early auditory experience on the development of local song preference in female swamp sparrows. Behav Ecol Sociobiol. 68:437–447. doi:10.1007/s00265-013-1658-7.

Anderson RC, Searcy WA, Peters S, Hughes M, DuBois AL, Nowicki S. 2017. Song learning and cognitive ability are not consistently related in a songbird. Anim Cogn. 20:309–320. doi:10.1007/s10071-016-1053-7.

Ashton BJ, Ridley AR, Edwards EK, Thornton A. 2018. Cognitive performance is linked to group size and affects fitness in Australian magpies. Nature. 554:364–367. doi:10.1038/nature25503.

Bioacoustics Research Program, Program BR. 2014. Raven Pro: Interactive Sound Analysis Software (Version 1.5). Ithaca, NY Cornell Lab Ornithol.

Boogert NJ, Anderson RC, Peters S, Searcy WA, Nowicki S. 2011. Song repertoire size in male song sparrows correlates with detour reaching, but not with other cognitive measures. Anim Behav. 81:1209–1216. doi:10.1016/j.anbehav.2011.03.004.

Boogert NJ, Fawcett TW, Lefebvre L. 2011. Mate choice for cognitive traits: A review of the evidence in nonhuman vertebrates. Behav Ecol. 22:447–459. doi:10.1093/beheco/arq173.

Boogert NJ, Giraldeau LA, Lefebvre L. 2008. Song complexity correlates with learning ability in zebra finch males. Anim Behav. 76:1735–1741. doi:10.1016/j.anbehav.2008.08.009.

Bouchard J, Goodyer W, Lefebvre L. 2007. Social learning and innovation are positively correlated in pigeons (Columba livia). Anim Cogn. 10:259–266. doi:10.1007/s10071-006-0064-1.

Buchanan KL, Leitner S, Spencer KA, Goldsmith AR, Catchpole CK. 2004. Developmental stress selectively affects the song control nucleus HVC in the zebra finch. Proc R Soc B Biol Sci. 271:2381–2386. doi:10.1098/rspb.2004.2874.

Buchanan KL, Spencer KA, Goldsmith AR, Catchpole CK. 2003. Song as an honest signal of past developmental stress in the European starling (Sturnus vulgaris). Proc R Soc London. 270:1149–1156. doi:10.1016/S0018-506X(03)00124-7.

Catchpole CK, Slater PJB. 2008. Bird Song: Biological Themes and Variations. 2nd ed. Cambridge: Cambridge University Press.

ten Cate C. 1985. Directed song of male zebra finches as a predictor of subsequent intra- and interspecific social behaviour and pair formation. Behav Processes. 10:369–374. doi:10.1016/0376-6357(85)90037-3.

Cauchard L, Boogert NJ, Lefebvre L, Dubois F, Doligez B. 2013. Problem-solving performance is correlated with reproductive success in a wild bird population. Anim Behav. 85:19–26. doi:10.1016/j.anbehav.2012.10.005.

Chantal V, Gibelli J, Dubois F. 2016. Male foraging efficiency, but not male problem-solving performance, influences female mating preferences in zebra finches. PeerJ. 4:e2409. doi:10.7717/peerj.2409.

Chen J, Zou Y, Sun Y-H, Cate C ten. 2019. Problem-solving males become more attractive to female budgerigars. Science (80-). 363:166–167. doi:10.1126/SCIENCE.AAU8181.

9

Cole EF, Morand-Ferron J, Hinks AE, Quinn JL. 2012. Cognitive ability influences reproductive life history variation in the wild. Curr Biol. 22:1808–1812. doi:10.1016/j.cub.2012.07.051.

Darwin C. 1871. The descent of man: and selection in relation to sex. London: J. Murray.

David P, Bjorksten T, Fowler K, Pomiankowski A. 2000. Condition-dependent signalling of genetic variation in stalk-eyed flies. Nature. 406:186–188. doi:10.1038/35018079.

Doucet SM, Montgomerie R. 2003. Multiple sexual ornaments in satin bowerbirds: Ultraviolet plumage and bowers signal different aspects of male quality. Behav Ecol. 14:503–509. doi:10.1093/beheco/arg035.

Dougherty LR, Guillette LM. 2018. Linking personality and cognition: a meta-analysis. Philos Trans R Soc B Biol Sci. 373:20170282. doi:10.1098/rstb.2017.0282.

DuBois AL, Nowicki S, Peters S, Rivera-Cáceres KD, Searcy WA. 2018. Song is not a reliable signal of general cognitive ability in a songbird. Anim Behav. 137:205–213. doi:10.1016/j.anbehav.2018.01.020.

Erhard HW, Boissy A, Rae MT, Rhind SM. 2004. Effects of prenatal undernutrition on emotional reactivity and cognitive flexibility in adult sheep. Behav Brain Res. 151:25–35. doi:10.1016/j.bbr.2003.08.003.

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. doi:10.1007/s10071-015-0908-7.

Fisher MO, Nager RG, Monaghan P. 2006. Compensatory growth impairs adult cognitive performance. PLoS Biol. 4:1462–1466. doi:10.1371/journal.pbio.0040251.

Fordyce JA, Gompert Z, Forister ML, Nice CC. 2011. A hierarchical bayesian approach to ecological count data: A flexible tool for ecologists. PLoS One. 6:1–7. doi:10.1371/journal.pone.0026785.

Griffin AS, Guillette LM, Healy SD. 2015. Cognition and personality: An analysis of an emerging field. Trends Ecol Evol. 30:207–214. doi:10.1016/j.tree.2015.01.012.

Hill GE. 1991. Plumage coloration is a sexually selected indicator of male quality. Nature. 350:337–339.

Holveck MJ, Riebel K. 2007. Preferred songs predict preferred males: consistency and repeatability of zebra finch females across three test contexts. Anim Behav. 74:297–309. doi:10.1016/j.anbehav.2006.08.016.

Holzer B, Jacot A, Brinkhof MWG. 2003. Condition-dependent signaling affects male sexual attractiveness in field crickets, Gryllus campestris. Behav Ecol. 14:353–359. doi:10.1093/beheco/14.3.353.

van Horik JO, Madden JR. 2016. A problem with problem solving: Motivational traits, but not cognition, predict success on novel operant foraging tasks. Anim Behav. 114:189–198. doi:10.1016/j.anbehav.2016.02.006.

Houde AE, Torio AJ. 1992. Effect of parasitic infection on male color pattern and female choice in guppies. Behav Ecol. 3:346–351. doi:1045-2249/92/.

Hume DK, Montgomerie R. 2001. Facial attractiveness signals different aspects of "quality" in women and men. Evol Hum Behav. 22:93–112. doi:10.1016/S1090-5138(00)00065-9.

Karino K, Utagawa T, Shinjo S. 2005. Heritability of the algal-foraging ability: An indirect benefit of female mate preference for males' carotenoid-based coloration in the guppy, Poecilia reticulata. Behav Ecol Sociobiol. 59:1–5. doi:10.1007/s00265-005-0001-3.

Keagy J, Savard JF, Borgia G. 2009. Male satin bowerbird problem-solving ability predicts mating success. Anim Behav. 78:809–817. doi:10.1016/j.anbehav.2009.07.011.

Kitaysky AS, Kitaiskaia E V, Piatt JF, Wingfield JC. 2006. A mechanistic link between chick diet and decline in seabirds? Proc R Soc B Biol Sci. 273:445–450. doi:10.1098/rspb.2005.3351.

Kriengwatana B, MacDougall-Shackleton SA, Farrell TM, Garcia L, Aitken SDT. 2015. Early-life nutritional stress affects associative learning and spatial memory but not performance on a novel object test. Behaviour. 152:195–218. doi:10.1163/1568539X-00003239.

Kroodsma DE. 1989. Suggested experimental designs for song playbacks. Anim Behav. 37:600–609. doi:10.1016/0003-3472(89)90039-0.

Lemon WC, Barth RH. 1992. The effects of feeding rate on reproductive success in the zebra finch, Taeniopygia guttata. Anim Behav. 44:851–857. doi:10.1016/S0003-3472(05)80581-0.

Lescroël A, Ballard G, Toniolo V, Barton KJ, Wilson PR, Lyver POB, Ainley DG. 2010. Working less to gain more: When breeding quality relates to foraging efficiency. Ecology. 91:2044–2055. doi:10.1890/09-0766.1.

Levitsky D, Strupp B. 1995. The Relationship between Undernutrition and Behavioral Development in Children Malnutrition and the Brain : Changing Concepts , Changing Concerns1. J Nutr. 125:2212–2220.

Little AC, Jones BC, DeBruine LM, Feinberg DR. 2008. Symmetry and sexual dimorphism in human faces: Interrelated preferences suggest both signal quality. Behav Ecol. 19:902–908. doi:10.1093/beheco/arn049.

MacKinlay RD, Shaw RC. 2018. Male New Zealand robin (Petroica longipes) song repertoire size does not correlate with cognitive performance in the wild. Intelligence.:0–1. doi:10.1016/j.intell.2018.10.009.

Medina-Garcia A, Jawor JM, Wright TF. 2017. Cognition, personality, and stress in budgerigars, Melopsittacus undulatus. Behav Ecol. 28:1504–1516. doi:10.1093/beheco/arx116.

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. doi:10.1016/j.anbehav.2015.07.002.

Neubauer RL. 1999. Super-normal length song preferences of female zebra finches (Taeniopygia guttata) and a theory of the evolution of bird song. Evol Ecol. 13:365–380. doi:10.1023/A:1006708826432.

Nowicki S, Searcy WA, Peters S. 2002. Brain development, song learning and mate choice in birds: A review and experimental test of the "nutritional stress hypothesis." J Comp Physiol A. 188:1003–1014. doi:10.1007/s00359-002-0361-3.

Orians GH. 1969. Age and hunting success in the brown pelican (Pelecanus occidentalis). Anim Behav. 17:316–319. doi:10.1016/0003-3472(69)90016-5.

Peters S, Searcy WA, Nowicki S. 2014. Developmental stress, song-learning, and cognition. Integr Comp Biol. 54:555–567. doi:10.1093/icb/icu020.

Plummer M, Best N, Cowles K, Vines K. 2006. CODA: Convergence Diagnosis and Output Analysis for MCMC. R News. 6:7–11.

Pravosudov V V., Lavenex P, Omanska A. 2005. Nutritional deficits during early development affect hippocampal structure and spatial memory later in life. Behav Neurosci. 119:1368–1374. doi:10.1037/0735-7044.119.5.1368.

Prior NH, Smith E, Lawson S, Ball GF, Dooling RJ. 2018. Acoustic fine structure may encode biologically relevant information for zebra finches. Sci Rep. 8:1–12. doi:10.1038/s41598-018-24307-0.

R Development Core Team. 2016. R: A Language and Environment for Statistical Computing. R Found Stat Comput Vienna Austria. 0:{ISBN} 3-900051-07-0. doi:10.1038/sj.hdy.6800737.

Riebel K. 2000. Early exposure leads to repeatable preferences for male song in female zebra finches. Proc R Soc B Biol Sci. 267:2553–2558. doi:10.1098/rspb.2000.1320.

Rowe C, Healy SD. 2014. Measuring variation in cognition. Behav Ecol. 25:1287–1292. doi:10.1093/beheco/aru090.

Santos de Souza A, da Camara Pacheco L, da Silva Castro P, Hokoc JN, Santos Rocha M, Tavares do Carmo M das G. 2008. Brain fatty acid profiles and spatial learning in malnourished rats: effects of nutritional intervention. Nutr Neurosci. 11:119–127. doi:10.1179/147683008X301504.

Searcy WA, Nowicki S. 2019. Birdsong learning, avian cognition, and the evolution of language. Anim Behav.:1–11. doi:10.1016/j.anbehav.2019.01.015.

Searcy WA, Peters S, Kipper S, Nowicki S. 2010. Female response to song reflects male developmental history in swamp sparrows. Behav Ecol Sociobiol. 64:1343–1349. doi:10.1007/S00265-0 10-0949-5.

Sewall KB, Soha J a, Peters S, Nowicki S. 2013. Potential trade-off between vocal ornamentation and spatial ability in a songbird. Biol Lett. 9:20130344. doi:10.1098/rsbl.2013.0344.

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. doi:10.1016/j.anbehav.2015.08.001.

Snowberg LK, Benkman CW. 2009. Mate choice based on a key ecological performance trait. J Evol Biol. 22:762–769. doi:10.1111/j.1420-9101.2009.01699.x.

Spencer KA, Buchanan KL, Goldsmith AR, Catchpole CK. 2003. Song as an honest signal of developmental stress in the zebra finch (Taeniopygia guttata). Horm Behav. 44:132–139. doi:10.1016/S0018-506X(03)00124-7.

Spencer KA, Wimpenny JH, Buchanan KL, Lovell PG, Goldsmith AR, Catchpole CK. 2005. Developmental stress affects the attractiveness of male song and female choice in the zebra finch (Taeniopygia guttata). Behav Ecol Sociobiol. 58:423–428. doi:10.1007/s00265-005-0927-5.

Stevenson JG. 1967. Reinforcing effects of chaffinch song. Anim Behav. 15:427–432. doi:10.1016/0003-3472(67)90040-1.

Tchernichovski O, Nottebohm F, Ho CE, Pesaran B, Mitra PP. 2000. A procedure for an automated measurement of song similarity. Anim Behav. 59:1167–1176. doi:10.1006/anbe.1999.1416.

Templeton CN, Laland KN, Boogert NJ. 2014. Does song complexity correlate with problem-solving performance inflocks of zebra finches? Anim Behav. 92:63–71. doi:10.1016/j.anbehav.2014.03.019.

Weathers WW, Sullivan KA. 1989. Juvenile Foraging Proficiency, Parental Effort, and Avian Reproductive Success. Ecol Monogr. 59:223–246.

Wilgers DJ, Hebets EA. 2015. Functional Approach to Condition. In: Irschick DJ, Briffa M, Podos J, editors. Animal Signaling and Function. Hoboken, New Jersey: John Wiley & Sons. p. 562–624.

Zann R. 1996. The zebra finch: a synthesis of field and laboratory studies. 5th ed. Oxford University Press.

Figure Legends:

Figure 1: Probability densities of conspecific preference estimates from hierarchical Bayesian statistical model for conspecific (black) vs. heterospecific (light grey) song. Solid lines indicate population-level estimates for conspecific song preference; dotted lines indicate individual-level estimates for conspecific song preference. Preference of 1.0 indicates 100% preference, which would mean that the bird hopped exclusively on perches triggering the given stimulus type. Preference of 0.0 indicates 0% preference, which would mean that the bird did not hop at all on the perches triggering the given stimulus type. Population preference for conspecific song was 0.705 (95% credible intervals: 0.687, 0.748) and population preference for heterospecific song was 0.295 (95% credible intervals: 0.252, 0.313). Model was run for 10,000 generations with a 1,000 generation burn-in.

Figure 2: Probability densities of Solver preference estimates from hierarchical Bayesian statistical model for Solver (black) vs. Non-Solver (light grey) song, where song preferences were aggregated for each female.

Solid lines indicate population-level estimates for Solver song preference; dotted lines indicate individual-level estimates for Solver song preference. Preference of 1.0 indicates 100% preference, which would mean that the bird hopped exclusively on perches triggering the given stimulus type. Preference of 0.0 indicates 0% preference, which would mean that the bird did not hop at all on the perches triggering the given stimulus type. Population preference for Solver song was 0.594 (95% credible intervals: 0.581, 0.628), and population preference for Non-Solver song was 0.406 (95% credible interval: 0.372, 0.419). Model was run for 10,000 generations with a 1,000 generation burn-in.

Figure 3: Probability densities of Solver song preference estimates from hierarchical Bayesian statistical model for Solver (black) vs. Non-Solver (light grey) song, where female preferences were aggregated for each song. Solid lines indicate group (Solver) level estimates; dotted lines indicate individual (stimulus song) level estimates. Preference of 1.0 indicates 100% preference, which would mean that the bird hopped exclusively on perches triggering the given stimulus type. Preference of 0.0 indicates 0% preference, which would mean that the bird did not hop at all on the perches triggering the given stimulus type. Aggregate preference for Solver song was 0.608 (95% confidence intervals: 0.592, 0.625). Aggregate preference for Non-Solver song was 0.392 (95% confidence intervals: 0.374, 0.408). Model was run for 10,000 generations with a 1,000 generation burn-in.
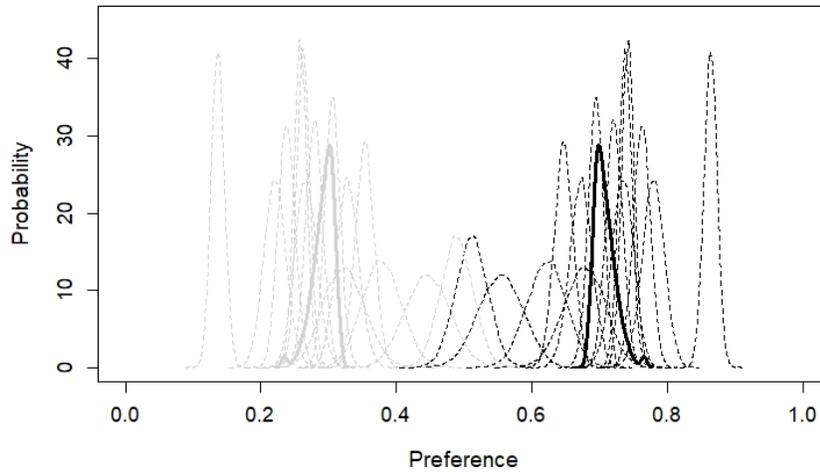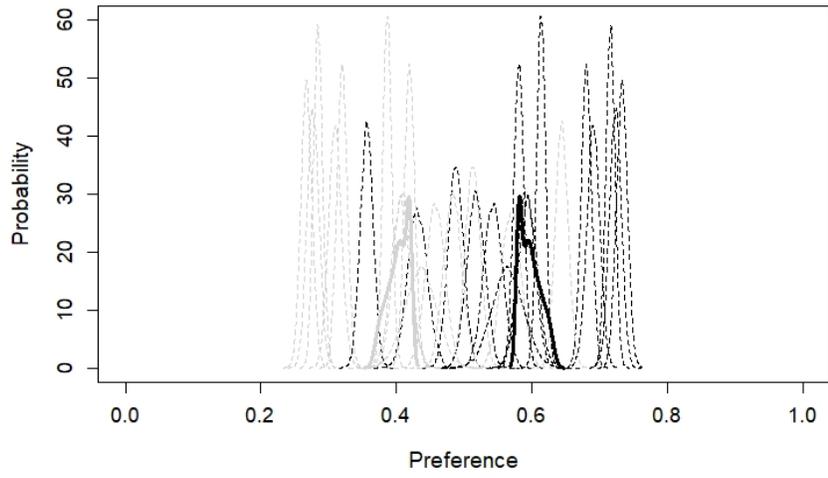
Figures:

Figure 1

Figure 2

Figure 3