Effects of auditory recognition learning on the perception of vocal features in European starlings (*Sturnus vulgaris*)

C. Daniel Meliza[a]

Department of Organismal Biology and Anatomy, University of Chicago, 1027 East 57th Street, Chicago, Illinois 60622

Learning to recognize complex sensory signals can change the way they are perceived. European starlings (*Sturnus vulgaris*) recognize other starlings by their song, which consists of a series of complex, stereotyped motifs. Song recognition learning is accompanied by plasticity in secondary auditory areas, suggesting that perceptual learning is involved. Here, to investigate whether perceptual learning can be observed behaviorally, a same–different operant task was used to measure how starlings perceived small differences in motif structure. Birds trained to recognize conspecific songs were better at detecting variations in motifs from the songs they learned, even though this variation was not directly necessary to learn the associative task. Discrimination also improved as the reference stimulus was repeated multiple times. Perception of the much larger differences between different motifs was unaffected by training. These results indicate that sensory representations of motifs are enhanced when starlings learn to recognize songs. © *2011 Acoustical Society of America*. [DOI: 10.1121/1.3641420]

I. INTRODUCTION

Vocalizations play an important role in the social behavior of songbirds. In many species, individual birds produce unique songs and identify each other by song (Stoddard, 1996), presumably by learning associations between specific acoustic features and the individuals that produce them (Beer, 1970). In order to form these associations, birds must be able to detect and discriminate the relevant acoustic features. Learning to recognize a song may modify sensory processing in order to facilitate recognition, a phenomenon broadly defined as perceptual learning (see Goldstone, 1998). To determine whether and how perceptions are modified by auditory learning of songs, European starlings (*Sturnus vulgaris*) were trained to recognize conspecific songs and then tested for their ability to discriminate between acoustic components of the songs.

Laboratory studies of auditory recognition behavior frequently rely on operant associational tasks in which subjects are presented with a stimulus, allowed to make one of several responses, and then rewarded or punished based on whether the response was the one the experimenter assigned to the stimulus (see Hall, 1991). A common design is the two-alternative-choice (2AC) task, where stimuli are associated with one of two distinct responses. The nature of the neuronal associations formed in the task can be probed by testing with novel stimuli related in some way to the training set. For example, starlings can be trained to associate songs with pecking behaviors in a 2AC task (Gentner and Hulse, 1998), and when trained subjects were presented with novel songs their responses were correlated with the proportion of familiar motifs in the new songs (Gentner and Hulse, 2000b). As shown in Fig. 1(a), motifs are temporally discrete elements of starling song (Adret-Hausberger and Jenkins,

1988; Eens *et al.,* 1989). The motifs in an individual's repertoire are generally unique to that individual (Eens, 1997), and the observation that starlings generalize on the basis of motifs indicates that in learning songs starlings form distinct memories of the component motifs and can recognize those motifs in novel contexts. Other studies indicate that information is also carried in the motif sequence and in sub-motif features (Gentner and Hulse, 1998; Gentner, 2008).

In these studies, it is not clear whether the formation of auditory stimulus-response associations involves any change to the processing and representation of the stimuli. Improvement on an associational task can come from perceptual learning or from the formation of stronger associations, or both, and the effects can be difficult to disambiguate (Hall, 1991; Weinberger, 2007). Perceptual ability needs to be tested independently of the specific response associations formed when the stimuli were learned. A useful paradigm for testing perception is the same–different (SD) task, which is not based on learned associations between responses and specific stimuli, but on perceived differences between pairs of stimuli (Park *et al.,* 1985; Brown and Dooling, 1987). Performance reflects the perceptual distinctiveness of the stimuli, and subjects can be tested with stimuli that they have never heard before or with stimuli heard in a different context. In this study, starlings were trained to recognize several conspecific songs in a 2AC task and then tested in a SD task for their ability to discriminate between the motif components of familiar and unfamiliar songs. The difference in performance on familiar and unfamiliar motifs was therefore an independent measure of the perceptual correlates of song familiarity.

II. GENERAL METHODS

A. Subjects

Twelve adult European starlings of both sexes (5 male, 5 female, 2 unknown) were captured from farms in northeastern

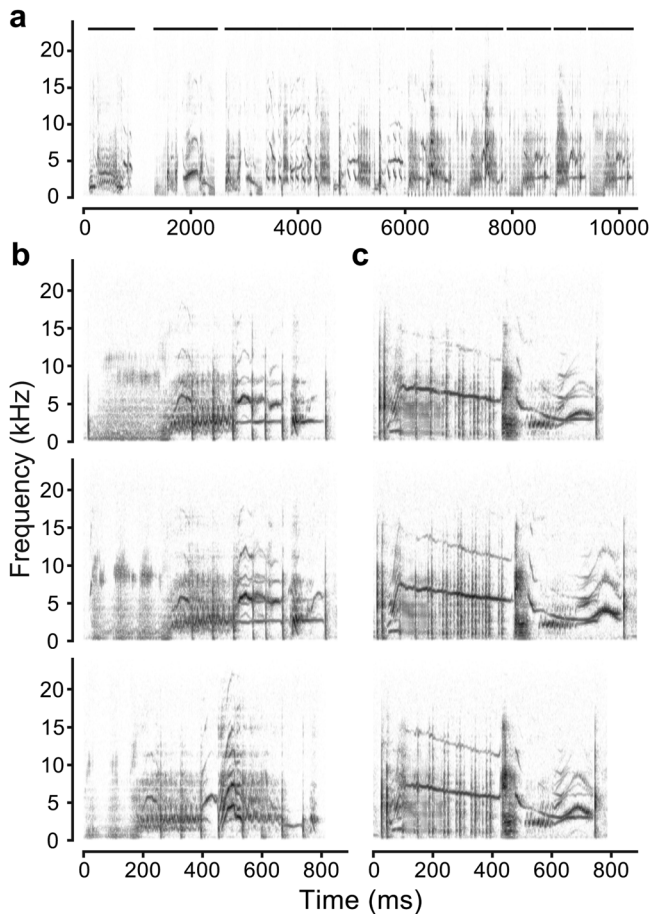[a] Author to whom correspondence should be addressed. Electronic mail: dmeliza@uchicago.edu

FIG. 1. Spectrograms of (a) a segment of a song bout from an adult male starling, with the temporal boundaries of motifs indicated by horizontal bars above, (b) the last motif in the song above (top) and two variants recorded in other songs, and (c) variants of a motif by a different singer. Variants may have differences in note composition [e.g., (b) top vs bottom] as well as differences in note features [e.g., compare duration of trilled note in the last half of (c) motifs]. The range of variation in motif structure shown here is typical, from slight differences in duration and timing of the component notes (c) to having less than 50% of their notes in common [(b), middle vs bottom].

Illinois. They were initially housed in mixed-sex flight aviaries where they received food and water *ad libitum*. At the beginning of the experiment, they were isolated in an operant apparatus and remained in isolation until all phases of training and testing were completed, with the exception of one bird who was housed with another bird, on free food, for 2 days between initial SD training and transfer. Birds were returned to the aviary at the completion of testing. The lighting schedule in the aviary and apparatus was matched to local daylight hours in Chicago. All animal procedures were performed according to protocols approved by the University of Chicago Institutional Animal Use and Care Committee and consistent with the guidelines of the National Institutes of Health. Sex was determined by visual examination of iris color, feather spot size, and bill color (Feare, 1984).

B. Stimuli

Song was recorded from three adult male starlings captured and housed under similar conditions as the experimental subjects. Some of the subjects were housed in the same aviary as some of the singers; the dates when birds occupied the aviary together were recorded and used to determine which singers and subjects were potentially socially familiar. During recording each bird was housed in isolation, in a 2 m^3 double-walled sound isolation booth (Industrial Acoustics Corporation, New York, NY). Recordings were made with an AT4071a directional microphone (Audio-Technica, Stow, OH) and amplified with a DMP3 microphone preamplifier (M-Audio, Irwindale, CA). The signals were digitized by a sound card with an integrated antialiasing filter, at 48 kHz and 16 bits/sample. Songs were stored to disk and digitally high-pass filtered (12 dB/octave) at 150 Hz and scaled to 96 dB peak amplitude. Between 100 and 300 complete song bouts were recorded from each bird over the course of several days.

From the song bouts recorded for each bird, 20 representative samples of 10 s each were extracted, sampling equally from the beginnings, middles, and ends of the bouts. Each sample (hereinafter song) was manually segmented into motifs (10–18 per sample; median 13) based on visual examination of the spectrograms. Motifs are temporally discrete vocal elements between 500 and 1 200 ms in length composed of a fairly stereotyped pattern of notes. Starlings typically sing the same motif two or three times in succession, though the structure varies somewhat between renditions and from bout to bout (Adret-Hausberger and Jenkins, 1988, Fig. 1). The features of individual notes vary as well as the note composition of the motif.

Motifs were segmented and categorized by the author and a second observer. Where segmentation boundaries were ambiguous, they were cross-checked against other bouts where the sequence of motifs or the spacing between motifs was different, with the goal of identifying consistent units of production. A total of 50, 38, and 48 unique motifs were identified from the songs of the three birds. The number of variants of each unique motif ranged from 1 to 32 (median 4). From each of the three birds, three variants of three motifs, for a total of 27 variants, were extracted using a 2 ms squared-cosine ramp at the temporal boundaries to eliminate transients.

A separate set of 15 motifs was extracted from an older library of starling songs (Meliza *et al.*, 2010). These stimuli were used only for initial training, so they were categorized more informally, and there was only one variant of each motif.

C. Apparatus

Starlings were trained and tested in a custom apparatus consisting of a galvanized wire-welded cage attached to a response panel. The entire apparatus was housed in an IAC-3 anechoic sound attention chamber (Industrial Acoustics, Bronx, NY). The response panel had three horizontally aligned response ports equipped with infrared emitters and detectors. When the birds probed these ports with their beaks, a natural foraging behavior for starlings (Feare, 1984), the light beam was interrupted and detected by a controller (ENV-253, Med Associates, St. Albans, VT). A light-emitting diode (LED) light was positioned at the back of each port.

The panel also housed two food hoppers operated by solenoids. Lighting was provided by a single 120 V incandescent bulb, controlled by a relay. A digital interface card (PCI-DIO96, Measurement Computing) controlled the solenoids and lighting and monitored the inputs from the response ports. Stimuli were presented through a small powered computer speaker (LCS-1040, Labtec, Singapore) positioned behind the response panels, out of the subjects' view. The analog signals for the speaker were produced by a multichannel sound card (Delta-1010, M-Audio). Stimulus presentation, response detection, and reinforcement were controlled with custom software written in C++. Birds were maintained in a closed economy in the apparatus and were allowed to initiate trials at any point during local daylight hours. Water was always available.

D. Training and testing

After being introduced into the apparatus, the birds were trained in an autoshaping procedure similar to the one described by Gentner and Hulse (1998). All the birds successfully learned to use the center port to initiate trials, and the right and left ports to report responses.

1. SD discrimination

Starlings were trained to respond to differences between motifs in an SD paradigm that required them to detect when a repeated background changed to a different motif (Park et al., 1985; Okanoya and Dooling, 1988). A schematic of the task is shown in Fig. 2(a). When the bird pecked the center port, a reference motif was presented repeatedly, separated by gaps of either 100 or 300 ms. During this response period (R1), which lasted up to ten repetitions, the LED in the center port flashed and the center port was active. If the bird failed to peck the center port during this period, the trial terminated without consequence. If the center port was pecked, a second response period (R2) was initiated after the motif currently playing ended. In R2, the center LED was extinguished, and the reference stimulus was repeated a random number of times, uniformly distributed between 1 and 7 (inclusive). After the last reference stimulus, a target stimulus was presented up to five times. During R2, the right port was monitored. If the bird pecked this port while the reference stimulus was still playing, the trial terminated at the end of the currently playing motif, and the lights in the sound attenuation chamber were extinguished for 10 s, during which no trials could be initiated. If the bird pecked after the transition to the target stimulus, playback ended and the bird was rewarded by access to the food hopper for 4 s. During training, target and reference stimuli were chosen randomly with replacement from a set of 9 motifs (1 variant each), which meant that in 1/9 of the trials the target stimulus was the same as the reference stimulus. During these catch trials, birds were punished if they made any response during R2. To facilitate learning, variants were not used during initial training so that all the stimulus pairs were highly dissimilar.

After each bird reached asymptotic performance, the stimulus set was expanded to a total of 15 motifs to test for generalization. Birds were also trained on sequences of two
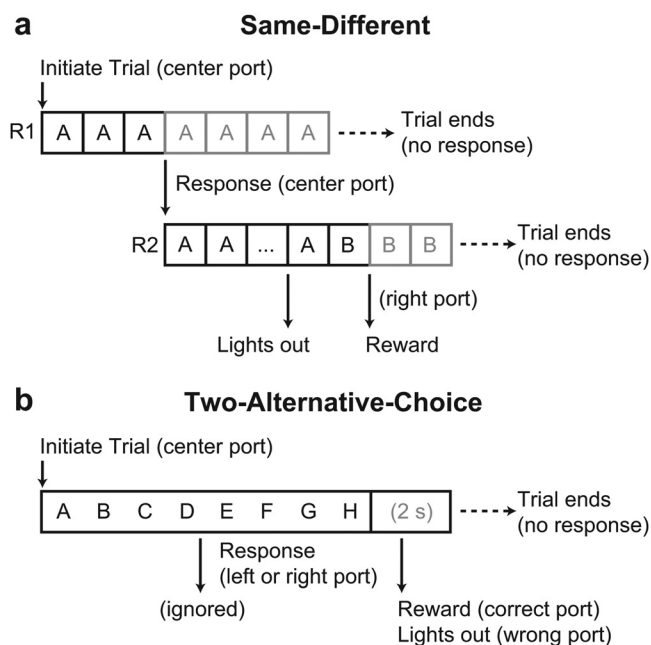

FIG. 2. Schematics describing the (a) repeated-background SD and (b) 2AC tasks. In the SD task (a), each trial consisted of two response intervals, R1 and R2. R1 was initiated when the bird pecked the center response port. The reference stimulus (A) played repeatedly until the bird pecked the center response port again, initiating R2. In R2, the reference stimulus repeated a random number of times and then transitioned to the target stimulus (B). Responses to a separate observation port were punished if they occurred before the transition, and rewarded if they occurred after. Failure to respond in R1 and R2 ended the trial without consequence. In the 2AC task (b), trials consisted of a stimulus playback period, initiated by a peck to the center port, followed by a 2 s silent response window. Responses were ignored during presentation of the stimulus, a randomly chosen song comprising a natural sequence of motifs (A–H). Each song was associated with either the left or right port; in the response interval the bird was rewarded or punished if it responded on the correct or incorrect port, and failure to respond ended the trial without consequence.

and three motifs to further encourage generalization and to prepare for a different set of experiments, not described here.

2. 2AC categorization

Following successful training on the SD task, the starlings were reshaped for a 2AC task. In this task, shown schematically in Fig. 2(b), starlings were trained to categorize 20 segments of starling song, each 10 s in length (Sec. II B). A peck to the center port initiated stimulus playback. After the stimulus completed playing, there was a 2 s response window during which the LEDs in both the left and right response ports flashed. Responses were rewarded by access to the food hopper, or punished by turning out the lights, depending on whether the bird correctly chose the port associated with the stimulus. If the bird failed to respond on either port during the response window, the trial terminated without reward or punishment.

Ten songs from one singer (Sec. II B) were assigned to one port, and ten to the other, which meant that each subject remained unexposed to the songs of one of the three singers. The port assignments and the identity of the bird whose songs were not presented were counterbalanced across the subjects. In each trial, the stimulus was chosen randomly with replacement from the set of 20 songs, except during correction trials,

which followed any incorrect response. In correction trials, which help to prevent the development of response biases, the stimulus was the same as the previous trial. Birds were trained on the 2AC task until they reached an average accuracy of at least 85% over three consecutive blocks of 100 trials. In some cases, subjects were first trained on the 2AC task using an unrelated set of stimuli from an older song library, and then transferred to the main set of 20 song segments.

3. Motif variant discrimination

After completing the 2AC training, birds were returned to the SD task, on the 15-motif training set (Sec. II D 1), until they returned to their previous level of performance. They were then switched to the test motifs (Sec. II B), consisting of three variants each of nine different motifs. The birds had encountered 6 of the motifs (18 variants total) in the 2AC training. The test procedure was identical to the SD training procedure, except that target and reference stimuli were chosen without replacement to ensure even sampling of all the reference–target pairs. If the bird failed to respond during R1, or responded too early in R2, so that it never heard the target stimulus, the same reference and target stimuli were used in the next trial. After all the pairs had been presented, the count was reset and another block of trials was presented, until each pair had been tested at least ten times.

The goal of the test procedure was to determine if familiarity with a motif significantly affected a bird's ability to discriminate between variants of that motif. In the 2AC training, each bird was required to categorize starling song containing six of the nine motifs subsequently tested in this phase. The 2AC training was balanced so that each of the test motifs was presented to some birds and not to others. The design was also balanced for social familiarity between subjects and singers.

E. Analysis

1. Training

Performance during training was assessed by grouping trials into blocks and calculating the average accuracy (2AC) or hit rate and false alarm rate (SD) in each block. Accuracy was defined as the proportion of trials in which the bird responded on the correct port, excluding correction trials and trials in which the bird made no response. The false alarm rate was the proportion of trials in which the bird responded in R2 before the transition to the target stimulus or, in the case of catch trials, at any point before the end of R2. Trials where the bird failed to respond in R1 or responded in R2 before the end of the first reference repetition were excluded. The hit rate was the proportion of trials in which the bird responded after the transition to the target stimulus, out of the subset of trials in which a transition occurred. The term "response rate" is used to refer to either false alarm or hit rate.

2. Testing

Performance on the SD task during the testing phase depended on a number of factors. Hit and false alarm rates varied between birds, presumably due to variation in individual ability and bias. Response rates also varied between stimuli, presumably reflecting differences in the similarity of (and difficulty in discriminating between) the motifs and variants. Response rates depended on the number of times the reference stimulus was presented before the target, and hit rates increased with the number of times the stimulus had been presented in the testing procedure (i.e., the block number). To estimate the effects of these factors, and of the bird's familiarity (social and operant) with the stimuli, the data were analyzed using a generalized linear mixed-effects model (GLMM). Stimulus (i.e., the ordered pair of motifs presented in a trial) and bird were modeled as random effects; that is, the effects associated with these factors were assumed to come from normal distributions with means of zero and variances σ_{bird}^2 and σ_{stim}^2. Fixed-effect factors were the number of reference motif repetitions, block number, social familiarity, and operant familiarity, as well as the first-order interactions. For trials where two different motifs were presented, a stimulus was considered familiar if either motif was. After it was determined that there was no main effect from social familiarity, the interactions involving this term were dropped from the model.

Hits and false alarms were fit separately, because they represented different kinds of responses. False alarms could occur on any trial, and when they did the bird did not hear the second stimulus. Hits could only occur on trials where the second stimulus was presented. The average response rate and the effect of many of the above-listed factors appeared to depend on whether the stimulus consisted of two different motifs, or two variants of the same motif (hereafter trial type). Therefore, in the hit model, the effect of trial type was modeled as an interaction with each of the fixed and random effects, so that there were separate estimates of each of the effects for the two trial types. The model also included a term, ρ_{bird}, that expressed the covariance, across birds, between performance on different trial types. Both hit and false alarm models were fit using MCMCGLMM (version 2.10; Hadfield, 2010), which computes confidence intervals (CIs) by drawing random samples from the posterior distribution of the model using Markov Chain Monte Carlo sampling. Priors were chosen to be non-informative on the response scale.

The parameter estimates from the GLMM are on a log odds scale, due to the logistic function linking the parameters to the binary response variable. A log odds of 0 corresponds to an odds ratio of 1:1, or 0.5 on the probability scale. Each unit on the scale corresponds to a change in the odds of a response by a factor of e. To conform with the model, several figures display average response rates on the logistic scale. Note that the logistic function is similar to the normal ogive, and thus the difference between hit rate and false alarm rate on the log odds scale is very close to the discriminability metric d' (Green and Swets, 1966). Confidence intervals for these graphs were estimated by non-parametric bootstrap sampling.

III. RESULTS

Starlings were trained to discriminate between motifs in an SD task, and reached a high level of accuracy. The time

C. Daniel Meliza: Auditory perceptual learning

course of acquisition is shown in Fig. 3(a): At first, both the false alarm and hit rate increased rapidly, followed by a slower decrease in the false alarm rate. At asymptote, the mean ($\pm$standard error, SEM) false alarm rate was 0.078 ± 0.016 and the hit rate was 0.86 ± 0.03. Once the birds had reached asymptotic behavior, the stimulus set was expanded from 9 to 15 motifs, greatly increasing the number of possible pairs. The hit rate in the first block after transfer decreased on average by 0.127 (± 0.04 SEM; $t_{11} = -3.07$, $p = 0.01$, paired t-test), but by the second block there was no significant difference with the asymptotic performance (mean $\pm$ SEM of the difference $= -0.047 \pm 0.03$; $t_{11} = -1.67$, $p = 0.12$).

Following SD training, the starlings were transferred to a 2AC task in which they were required to associate 10 s segments of song with responses on either the left or the right side of the response panel. Acquisition in this task is shown in Fig. 3(b). All but one of the birds acquired the new task easily; their average accuracy exceeded and remained above chance for at least 300 trials within 300–1 400 trials (median 800). The remaining bird took much longer to learn the task (2 600 trials) but eventually reached the same level of performance as the other birds. At asymptote, the mean ($\pm$SEM) accuracy was 0.917 ± 0.014.

After a brief reintroduction to the SD task using the training stimuli, the starlings were tested on their ability to discriminate between motif stimuli (three variants of nine different motifs). The birds had encountered six of the motifs during training. All of the possible pairs of the 27 stimuli were tested; 27 pairs were catch trials in which no transition occurred, 54 were comparisons between variants of the same motif, and 648 were comparisons between different motifs. During testing, the false alarm rate was not detectably

different from baseline (mean $\pm$ SEM of the difference $= -0.027 \pm 0.016$; $t_{11} = -1.6$, $p = 0.14$, paired t-test), nor was the response rate for comparisons between different motifs (0.006 ± 0.023; $t_{11} = 0.29$, $p = 0.78$). The mean hit rate ($\pm$SEM) for variants, which were not presented during training, was 0.346 ± 0.026. This value was significantly greater than the false alarm rate ($t_{11} = 9.4$, $p = 1.3 \times 10^{-6}$, paired t-test), and significantly less than the hit rate for motifs ($t_{11} = -18.5$, $p = 1.2 \times 10^{-9}$). Reduced discrimination between variants was also reflected in the reaction time (RT). Median RT for variants was 0.90 s, versus 0.82 s for motifs ($p < 1 \times 10^{-16}$, Wilcoxon rank-sum test), and the difference in mean RT was even greater (1.27 vs 0.98 s).

The effect of familiarity on discrimination is shown in Fig. 4. Response rates were greater for familiar than unfamiliar variants, with the largest effect observed when the reference motif had only been presented once. No such effect was observed for comparisons between different motifs. Based on a GLMM analysis, the effect of familiarity was significant, with an effect size indicating birds were 1.67 times as likely to detect the difference between motif variants when the motifs were familiar (Table I, stimulus familiar). There was also a significant negative interaction between familiarity and reference count, consistent with the decreasing effect of familiarity when the reference was repeated several times. At seven repetitions, the net effect of familiarity was negative, though not significantly (log odds -0.14; 95% CI -0.46–0.18). There was no detectable effect of familiarity on discrimination between motifs (see Table I), or on the false alarm rate (log odds 0.11; 95% CI -0.03–0.27). No effect of social familiarity was observed.
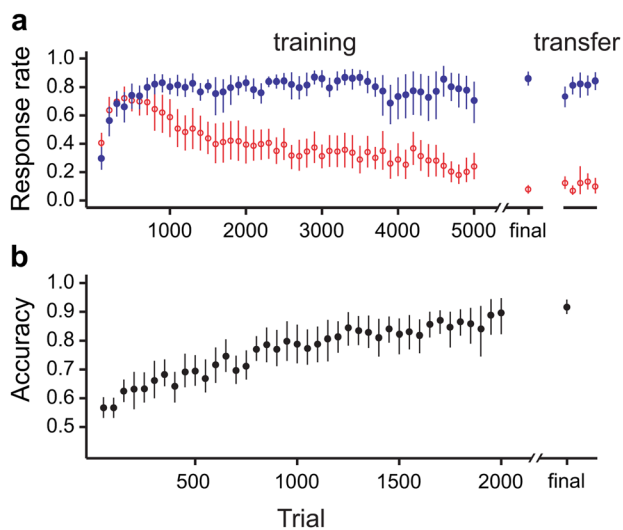

FIG. 3. (Color online) Acquisition curves for (a) SD and (b) 2AC tasks. In (a), the left panel shows initial acquisition of the SD task. Trials have been grouped into blocks of 100. Open circles show the proportion of false alarms and closed circles show the proportion of trials in which the transition to a different motif was correctly detected. Performance in the last 200 trials is shown by the final set of symbols. The right panel shows responses after transfer to a larger stimulus set. In (b), trials have been grouped into blocks of 50, and the average accuracy in each block is shown, with performance in the final 200 trials indicated by the final symbol. In both plots, symbols show mean across birds, with 95% confidence intervals estimated by bootstrapping.
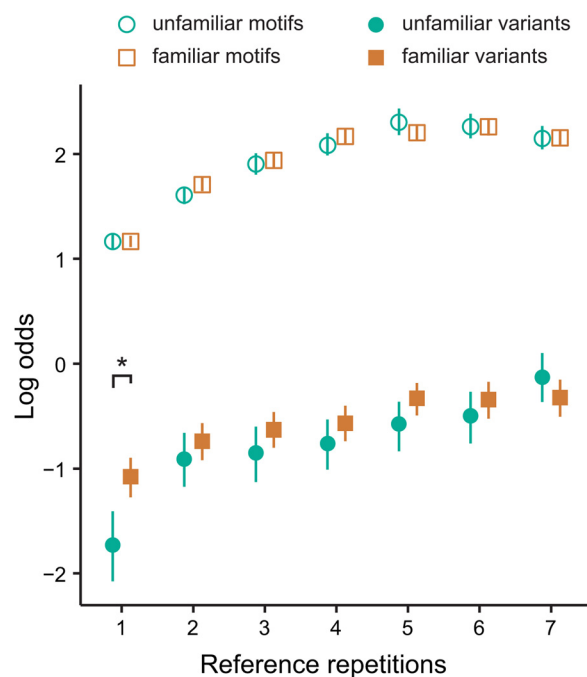

FIG. 4. (Color online) Hit rate for variants (closed symbols) and motifs (open symbols) as a function of familiarity and number of reference repetitions. Circles indicate mean response rate (log odds) for unfamiliar stimuli, and squares indicate mean response rate for familiar stimuli. Error bars are 95% confidence intervals, and the asterisk denotes a significant *post hoc* difference between familiar and unfamiliar ($p < 0.05$, GLMM).

TABLE I. Effect estimates (log odds), with lower and upper 95% confidence intervals, from GLMM analysis of SD discrimination performance. Effects for each type of trial (different motifs and different variants) are shown in separate columns.[a] In the interaction terms (e.g., fam × nref), fam refers to stimulus familiarity.

	Motifs		Variants	
μ[b]	0.70	[0.33, 1.05]*	−2.47	[−3.04, −1.91]*
Stimulus familiar[c]	−0.05	[−0.26, 0.14]	0.51	[0.12, 0.90]*
Social familiar[d]	0.02	[−0.05, 0.08]	−0.11	[−0.33, 0.10]
nref[e]	0.29	[0.25, 0.33]*	0.31	[0.23, 0.39]*
block[f]	0.09	[0.07, 0.12]*	0.11	[0.06, 0.17]*
fam × nref	−0.02	[−0.06, 0.02]	−0.09	[−0.15, −0.03]*
fam × block	0.04	[0.02, 0.07]*	0.00	[−0.04, 0.04]
nref × block	−0.02	[−0.02, −0.02]*	−0.01	[−0.02, 0.00]
σ_{stim}[g]	0.35	[0.32, 0.38]	0.88	[0.71, 1.06]
σ_{bird}[h]	0.49	[0.30, 0.74]	0.52	[0.29, 0.79]

[a]Fixed effects significantly greater or less than zero are indicated by an asterisk.
[b]Baseline probability of a response.
[c]Effect of exposure to stimuli during 2AC training.
[d]Effect of social exposure to the birds that sang the stimuli.
[e]Effect of each additional repetition of the reference stimulus.
[f]Testing block, or the number of times the bird had been exposed to a stimulus pair.
[g]Standard deviation of response rate across stimuli.
[h]Standard deviation of response rate across bird.

Response rates were also affected by a number of factors unrelated to familiarity, each shown in a panel of Fig. 5. Hit rate increased with the number of times the reference stimulus was repeated before the transition (Table I, nref). As shown in Fig. 5(a), for comparisons between variants, hit rate continued to increase with each additional repetition, but for comparisons between motifs the effect was more non-linear, reaching a plateau around the fourth or fifth repetition. Hit rate also increased gradually over the course of testing [Fig. 5(b); Table I, block].

As shown in Fig. 5(c), there was considerable variability in response rates across subjects, which was modeled as a random effect in the GLMM analysis (Table I, σ_{bird}). No difference was observed in between-subject variability for motif and variant trials ($p = 0.82$, GLMM), and the covariance was not significantly different from zero ($\rho_{bird} = -0.003$, $p = 0.97$, GLMM), indicating that performance on one type of trial was not predictive of a bird's performance on the other type. Between-subject variability might reflect differences in ability or simply in bias (i.e., a general tendency to respond regardless of the stimulus). However, the false alarm rate was negatively correlated with the hit rate for motifs ($r_{10} = -0.74$, $p = 0.006$, Pearson product-moment correlation) and not significantly correlated with the hit rate for variants ($r_{10} = -0.35$, $p = 0.27$), which is more consistent with differences in ability. Figure 5(d) shows the variability in hit rate across stimuli, which was also modeled as a random effect (Table I, σ_{stim}). Response rates varied much more across variants than across motifs ($p = 0.0003$, GLMM), indicating that some variants were much easier to tell apart than others. The difference in average response rate between trial types was much larger than the variance within each type (Cohen's $d = 3.36$; 95% CI 2.5–4.3).
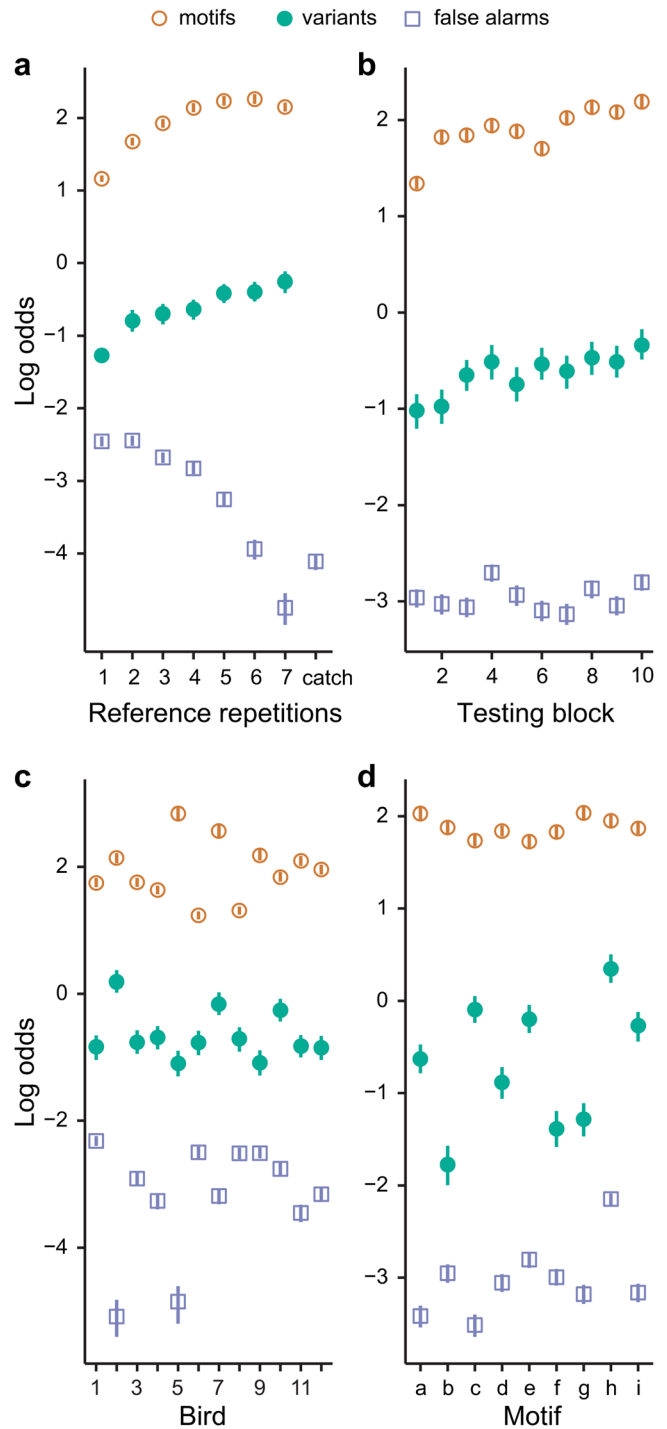

FIG. 5. (Color online) Response rates (log odds scale) as a function of (a) the number of times the reference stimulus was repeated, (b) the testing block, or number of times the bird had heard a given stimulus pair, (c) the subject, and (d) the motif. Open circles indicate the mean hit rate for trials where the stimuli were different motifs; closed circles indicate the hit rate for trials where the stimuli were different variants of the same motif; and open squares indicate the false alarm rate for all trials. Error bars indicate 95% confidence intervals. Catch trials, in which the reference stimulus was repeated more than seven times, and no transition to a target occurred, are grouped into a single category in (a); because the reference motif could repeat up to five times during this interval, the false alarm rate is corrected to reflect the average probability of a response during one of the repetitions. In (d), the hit rate is shown as a function of the reference motif. The variants of motif "b" are shown in Fig. 1(c), and the variants of "e" in Fig. 1(b).

C. Daniel Meliza: Auditory perceptual learning

IV. DISCUSSION

Starlings were trained to recognize the songs of individual conspecifics in a 2AC operant task. Each subject learned to categorize songs from two different birds. They were then tested in a separate SD task on their ability to discriminate between motifs and variants from these songs, and from songs that had not been encountered in training. The birds showed an increased ability to discriminate between variants of familiar motifs. This effect was greatest when the reference motif was only played once, and was not observed when birds were comparing different motifs.

Several other factors also influenced performance. Different motifs were considerably easier to discriminate than variants of the same motif. Performance increased over the course of testing, presumably as the birds became more familiar with the stimuli. Discrimination also improved as the reference stimulus was repeated multiple times. The false alarm rate decreased as hit rates increased, indicating a true increase in performance, rather than increasing difficulty in withholding responses or a strategy based on waiting until late in the trial when the probability of a transition having occurred was higher. A similar "perceptual anchoring" effect has been observed in numerous human psychophysical studies (see Braida *et al.*, 1984), as well as in starlings trained in a SD paradigm to discriminate between tones (Zokoll *et al.*, 2007).

A. Effects of auditory learning on motif perception

Perceptual learning implies that sensory stimuli are perceived differently across a range of tasks and behavioral contexts (Hall, 1991; Goldstone *et al.*, 2001). This study demonstrates auditory perceptual learning in European starlings by showing that familiarity with songs learned in one operant task was correlated with increased discrimination of the component motifs in a separate task, an acquired distinctiveness (Lawrence, 1949). Interestingly, the acoustic differences that became more distinct were not relevant to the original 2AC task. All the variants of each motif were associated with the same port, which meant that the differences between variants carried no information about the correct choice in the 2AC task. Though surprising, this result makes it especially unlikely that improved performance in the SD task reflects some aspect of the associational learning carried over from the 2AC task. Instead, increased discrimination ability would appear to be the result of a separate process occurring at the level of sensory representation.

In human psychophysical studies, tasks that require subjects to make categorical distinctions between sensory stimuli tend to result in better discrimination of features that distinguish categories (e.g., Lawrence, 1949; Goldstone *et al.*, 2001), sometimes with a concomitant reduction in the perceived difference between exemplars within the same category (Goldstone, 1994; Guenther *et al.*, 1999). Based on these findings, one might have predicted that starlings would show increased discrimination between motifs and reduced discrimination between variants, but the opposite obtained. Under some conditions perceptual learning can affect task-irrelevant features (Watanabe *et al.*, 2001; Amitay *et al.*, 2006), particularly when the training task is easy (Ahissar

and Hochstein, 1997). The present results may therefore reflect the easiness of the operant task relative to the typical acoustic conditions in which a starling would encounter new songs. Noisy environments like those found in a starling flock might restrict perceptual learning to the most robust features or drive increased tolerance to noise (Li and DiCarlo, 2008). Seeba and Klump (2009) found that social familiarity led to increased perceptual tolerance to noise or gaps inserted into motifs. The present study tested discrimination, not tolerance, which may explain why there was no significant effect from social familiarity.

A related question is why no effect of familiarity was observed on discrimination between different motifs, even though the increased performance on variants indicates that the sensory representations of the motifs had changed. The changes may have been subtle and may only have been observable in the most difficult trials, when the bird was comparing variants and only heard the reference motif once. This hypothesis is supported by the negative interaction between familiarity and reference count (see Fig. 4). However, there was no detectable negative interaction between familiarity and testing block. Even though the odds of a correct response increased by a factor of more than 1.8 over the course of training, familiarity continued to have a positive effect on performance. Thus, a different form of learning may have contributed to improvement over the course of testing, and one possibility is that comparisons may take place at multiple levels of the auditory pathway, not all equally affected by the song recognition learning. Response latencies were considerably shorter for motif comparisons, which could imply that the easier comparisons are based on fairly high-level representations, whereas more difficult comparisons recruit lower-level representations (Ahissar *et al.*, 2009). If this is the case, the present results could suggest that perceptual learning in the 2AC task took place at these lower levels, having little effect on comparisons between different motifs and leaving room for further learning during the testing.

B. Individual and stimulus variation

One of the major strengths of mixed-effects models is the ability to estimate variance due to uncontrolled factors. In this study, these factors were individual ability and stimulus discriminability. Individual variation reflected differences in discrimination performance, rather than bias, because average hit rates were not positively correlated with false alarm rates. Variation was similar for comparisons between motifs and variants, and there was no significant correlation between hit rates on the two kinds of trials. A positive correlation would imply that performance on both trial types was related, and a negative correlation would imply that birds tended to be better at one task or the other. Neither hypothesis is supported by the data.

Performance also varied between stimuli. The largest contributing factor was whether the stimulus consisted of two different motifs, or whether the motifs were variants of each other. The difference in means between these groups was much larger than the variance within them, indicating that motifs identified by spectrographic examination as

variants of each other were in fact perceptually similar to the starlings. Variation in performance across stimuli in the motif trials, however, was much lower than in the variant trials, indicating that some variants were quite similar to each other, and others were much more distinct. Based on informal comparisons of motif spectrograms, performance appeared to reflect the number of notes shared between variants. For example, the variants of the motif shown in Fig. 1(c) are nearly identical, and the average hit rate was correspondingly low [see Fig. 5(d), motif b]. In contrast, the variants of the motif in Fig. 1(b) are more dissimilar, and the average hit rate was much higher [Fig. 5(d), motif e]. Further work, involving systematic manipulations of the note contents of each motif, would be necessary to confirm whether note sharing or any other measure of similarity corresponds to starling perceptions.

The hit rate for variants was greater than the false alarm rate even though the birds were not initially trained to discriminate between variants. Thus, although starlings primarily use motifs and sequences of motifs in their recognition behavior (Gentner, 2008), they remain capable of perceiving differences between motifs at a much finer scale (cf. Lohr et al., 2006). There may be some parallels to the ability of humans to attend separately to structural and transformational variants of speech. Fine-scale variation in starling song may reflect constraints in production or processing, or it may carry ethologically relevant signals, perhaps about identity, emotional state, or spatial location. Female starlings exhibit preferences among male songs after hearing only a few motifs (Gentner and Hulse, 2000a), a behavior that might depend on fine-scale structure.

V. CONCLUSIONS

The European starling is an adept auditory learner, and there is a growing body of work examining this ability from behavioral and neurobiological perspectives (e.g., Leppelsack and Vogt, 1976; Müller and Leppelsack, 1985; Gentner and Hulse, 1998, 2000b; Meliza et al., 2010). The combination of these approaches promises to yield important insights into how auditory processing circuits can robustly extract behaviorally salient signals from the environment while remaining flexible enough to adapt to new situations. Recent work has shown that learning to recognize songs involves extensive neuronal plasticity in secondary auditory areas (Gentner and Margoliash, 2003; Thompson and Gentner, 2010), but the perceptual consequences of this plasticity have been unknown. This study demonstrates that when starlings learn to recognize song in an operant categorization task, perceptual learning occurs, as evidenced by an increased ability to discriminate between closely similar variants of motifs from those songs. The specificity of the effect for variants may indicate that the underlying changes to sensory processing occur at a fairly early stage in the auditory pathway.

Adret-Hausberger, M., and Jenkins, P. (**1988**). "Complex organization of the warbling song in starlings," Behaviour **107**, 138–156.

Ahissar, M., and Hochstein, S. (**1997**). "Task difficulty and the specificity of perceptual learning" Nature (London) **387**, 401–406.

Ahissar, M., Nahum, M., Nelken, I., and Hochstein, S. (**2009**). "Reverse hierarchies and sensory learning," Philos. Trans. R. Soc. London, Ser. B **364**, 285–299.

Amitay, S., Irwin, A., and Moore, D. R. (**2006**). "Discrimination learning induced by training with identical stimuli," Nat. Neurosci. **9**, 1446–1448.

Beer, C. (**1970**). "Individual recognition of voice in the social behavior of birds," Adv. Study Behav. **3**, 27–74.

Braida, L. D., Lim, J. S., Berliner, J. E., Durlach, N. I., Rabinowitz, W. M., and Purks, S. R. (**1984**). "Intensity perception. XIII. Perceptual anchor model of context-coding," J. Acoust. Soc. Am. **76**, 722–731.

Brown, S. D., and Dooling, R. J. (**1987**). "Perception of bird calls by budgerigars and humans: A multidimensional scaling analysis," J. Acoust. Soc. Am. **81**, S34–S34.

Eens, M. (**1997**). "Understanding the complex song of the European starling: An integrative approach," Adv. Study Behav. **26**, 355–434.

Eens, M., Pinxten, R., and Verheyen, R. F. (**1989**). "Temporal and sequential organization of song bouts in the European starling," Ardea **77**, 75–86.

Feare, C. (**1984**). *The Starling* (Oxford University Press, Oxford), pp. 6, 24.

Gentner, T. Q. (**2008**). "Temporal scales of auditory objects underlying birdsong vocal recognition," J. Acoust. Soc. Am. **124**, 1350–1359.

Gentner, T. Q., and Hulse, S. H. (**1998**). "Perceptual mechanisms for individual vocal recognition in European starlings, Sturnus vulgaris," Anim. Behav. **56**, 579–594.

Gentner, T. Q., and Hulse, S. H. (**2000a**). "Female European starling preference and choice for variation in conspecific male song," Anim. Behav. **59**, 443–458.

Gentner, T. Q., and Hulse, S. H. (**2000b**). "Perceptual classification based on the component structure of song in European starlings," J. Acoust. Soc. Am. **107**, 3369–3381.

Gentner, T. Q., and Margoliash, D. (**2003**). "Neuronal populations and single cells representing learned auditory objects," Nature (London) **424**, 669–674.

Goldstone, R. L. (**1994**). "Influences of categorization on perceptual discrimination," J. Exp. Psychol. Gen. **123**, 178–200.

Goldstone, R. L. (**1998**). "Perceptual learning," Annu. Rev. Psychol. **49**, 585–612.

Goldstone, R. L., Lippa, Y., and Shiffrin, R. M. (**2001**). "Altering object representations through category learning," Cognition **78**, 27–43.

Green, D., and Swets, J. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York), Chaps. 2–3.

Guenther, F. H., Husain, F. T., Cohen, M. A., and Shinn-Cunningham, B. G. (**1999**). "Effects of categorization and discrimination training on auditory perceptual space," J. Acoust. Soc. Am. **106**, 2900–2012.

Hadfield, J. (**2010**). "MCMC methods for multi-response generalised linear mixed models: The MCMCglmm R package," J. Stat. Software **33**, 1–22.

Hall, G. (**1991**). *Perceptual and Associative Learning* (Oxford University Press, Oxford), Chap. 1, pp. 1–28.

Lawrence, D. (**1949**). "Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus," J. Exp. Pyschol. **39**, 770–784.

Leppelsack, H.-J., and Vogt, M. (**1976**). "Responses of auditory neurons in the forebrain of a songbird to stimulation with species-specific sounds," J. Comp. Physiol., A **107**, 263–274.

Li, N., and DiCarlo, J. J. (**2008**). "Unsupervised natural experience rapidly alters invariant object representation in visual cortex," Science **321**, 1502–1507.

Lohr, B., Dooling, R. J., and Bartone, S. (**2006**). "The discrimination of temporal fine structure in call-like harmonic sounds by birds," J. Comp. Psychol. **120**, 239–251.

Meliza, C. D., Chi, Z., and Margoliash, D. (**2010**). "Representations of conspecific song by starling secondary forebrain auditory neurons: Towards a hierarchical framework," J. Neurophysiol. **103**, 1195–1208.

Müller, C. M., and Leppelsack, H.-J. (**1985**). "Feature extraction and tonotopic organization in the avian auditory forebrain," Exp. Brain Res. **59**, 587–599.

Okanoya, K., and Dooling, R. J. (**1988**). "Obtaining acoustic similarity measures from animals: A method for species comparisons," J. Acoust. Soc. Am. **83**, 1690–1693.

Park, T., Okanoya, K., and Dooling, R. J. (**1985**). "Operant conditioning of small birds for acoustic discrimination," Ethology **3**, 5–9.

Seeba, F., and Klump, G. M. (**2009**). "Stimulus familiarity affects perceptual restoration in the European starling (Sturnus vulgaris)," PLoS ONE 4, e5974.

Stoddard, P. K. (**1996**). "Vocal recognition of neighbors by territorial passerines," in Ecology and Evolution of Acoustic Communication in Birds, edited by D. E. Kroodsma and E. H. Miller (Cornell University Press, Ithaca, NY), pp. 356–374.

Thompson, J. V., and Gentner, T. Q. (**2010**). "Song recognition learning and stimulus-specific weakening of neural responses in the avian auditory forebrain," J. Neurophysiol. **103**, 1785–1797.

Watanabe, T., Náñez, J. E., and Sasaki, Y. (**2001**). "Perceptual learning without perception," Nature (London) **413**, 844–848.

Weinberger, N. M. (**2007**). "Auditory associative memory and representational plasticity in the primary auditory cortex," Hear. Res. **229**, 54–68.

Zokoll, M. A., Klump, G. M., and Langemann, U. (**2007**). "Auditory short-term memory persistence for tonal signals in a songbird," J. Acoust. Soc. Am. **121**, 2842–2851.