Title

Lesions to Caudomedial Nidopallium Impair Individual Vocal Recognition in the Zebra Finch

Journal

Journal of Neuroscience, 43(14)

ISSN

0270-6474

Authors

Yu, Kevin

Wood, William E

Johnston, Leah G

et al.

Publication Date

2023-04-05

DOI

10.1523/jneurosci.0643-22.2023

Peer reviewed

Behavioral/Cognitive

Lesions to Caudomedial Nidopallium Impair Individual Vocal Recognition in the Zebra Finch

Kevin Yu,[1*] **William E. Wood,**[1*] **Leah G. Johnston,**[2] **and Frederic E. Theunissen**[1,3,4]

[1]Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley California 94720, [2]Herbert Wertheim School of Optometry and Vision Science, University of California, Berkeley, Berkeley California 94720, [3]Departments of Psychology, and [4]Integrative Biology, University of California, Berkeley, Berkeley California 94720

Many social animals can recognize other individuals by their vocalizations. This requires a memory system capable of mapping incoming acoustic signals to one of many known individuals. Using the zebra finch, a social songbird that uses songs and distance calls to communicate individual identity (Elie and Theunissen, 2018), we tested the role of two cortical-like brain regions in a vocal recognition task. We found that the rostral region of the Cadomedial Nidopallium (NCM), a secondary auditory region of the avian pallium, was necessary for maintaining auditory memories for conspecific vocalizations in both male and female birds, whereas HVC (used as a proper name), a premotor areas that gates auditory input into the vocal motor and song learning pathways in male birds (Roberts and Mooney, 2013), was not. Both NCM and HVC have previously been implicated for processing the tutor song in the context of song learning (Sakata and Yazaki-Sugiyama, 2020). Our results suggest that NCM might not only store songs as templates for future vocal imitation but also songs and calls for perceptual discrimination of vocalizers in both male and female birds. NCM could therefore operate as a site for auditory memories for vocalizations used in various facets of communication. We also observed that new auditory memories could be acquired without intact HVC or NCM but that for these new memories NCM lesions caused deficits in either memory capacity or auditory discrimination. These results suggest that the high-capacity memory functions of the avian pallial auditory system depend on NCM.

Key words: **auditory cortex; auditory memory; auditory pallium; individual recognition; voice neuron; voice recognition**

Significance Statement

Many aspects of vocal communication require the formation of auditory memories. Voice recognition, for example, requires a memory for vocalizers to identify acoustical features. In both birds and primates, the locus and neural correlates of these high-level memories remain poorly described. Previous work suggests that this memory formation is mediated by high-level sensory areas, not traditional memory areas such as the hippocampus. Using lesion experiments, we show that one secondary auditory brain region in songbirds that had previously been implicated in storing song memories for vocal imitation is also implicated in storing vocal memories for individual recognition. The role of the neural circuits in this region in interpreting the meaning of communication calls should be investigated in the future.

Introduction

Successful vocal interactions often require individuals to recognize the identity of another vocalizer. This vocal-based individual recognition requires the brain to store memories of known individuals and to map the acoustic features of a sound to one of potentially hundreds of known individuals (Tibbetts and Dale, 2007; Carlson et al., 2020). The zebra finch is a social songbird that uses at least two vocalization types among its repertoire, the song and distance call (DC), to signal vocalizer identity (Zann, 1996; Elie and Theunissen, 2016). These two call types have acoustic signatures that are unique to each individual and stereotyped within an individual (D'Amelio et al., 2017; Elie and Theunissen, 2018). We have previously shown that zebra finches have a large-capacity memory for recognizing conspecific vocalizers based on

Received Mar. 31, 2022; revised Feb. 20, 2023; accepted Feb. 23, 2023.

Author contributions: K.Y., W.E.W., and F.E.T. designed research; K.Y., W.E.W., and L.G.J. performed research; K.Y., W.E.W., F.E.T., and L.G.J. analyzed data; K.Y. and F.E.T. wrote the paper.

This research was supported by National Institute on Deafness and Other Communication Disorders Grant R01 018321 to F.E.T. and National Science Foundation Graduate Fellowship DGE 1752814 to K.Y. and a National Institute of Health Graduate Training Grant T32EY007043 to L.G.J. We thank D. Perkel and A. Leblois for contributing to the song stimuli used in the behavioral task; I. Rice, A. Prasad, R. Vu, and C. Chen for animal training and behavioral testing; the Cancer Research Laboratory Molecular Imaging Center, which is supported by University of California, Berkeley Biological Faculty Research Fund Histological for slice imaging and digitizing; and Holly Aaron and Feather Ives for microscopy advice and support.

*K.Y. and W.E.W. contributed equally to this work.

The authors declare no competing financial interests.

Correspondence should be addressed to Frederic E. Theunissen at theunissen@berkeley.edu.

https://doi.org/10.1523/JNEUROSCI.0643-22.2023

their acoustic signatures and that those auditory memories are learned quickly and can persist for several weeks (Yu et al., 2020). How and where are these memories formed, stored, and retrieved in the brain?

Studies in humans and primates have shown that sensory cortical regions, including auditory association cortex, are involved in the memory and classification of behaviorally relevant sounds (Weinberger, 2004; Perrodin et al., 2011), including features required for voice recognition in humans (Formisano et al., 2008). One region of the avian forebrain thought to be analogous to auditory association cortex is NCM (caudomedial nidopallium; Bolhuis and Gahr, 2006; Bolhuis et al., 2010). NCM has been indirectly implicated in memory and recognition of the tutor song through immediate early gene studies (Mello et al., 1995; Bolhuis et al., 2001) and stimulus-specific habituation (Chew et al., 1995) in adult birds, and changes in neural tuning (Phan et al., 2006; Yanagihara and Yazaki-Sugiyama, 2016) and pharmacological inactivations (London and Clayton, 2008; Pagliaro et al., 2020) in juvenile birds. In more direct evidence, pharmacological disruption of NCM during a learning task using pure tones reduces learning rate but not final performance, suggesting a role in association learning but not memory retrieval (Macedo-Lima and Remage-Healey, 2020). Lesions to NCM, in contrast to the manipulations cited above, have not been shown to affect song imitation learning (Canopoli et al., 2014, 2016, 2017), although lesions do impair song recognition when assayed with a tutor song preference test (Gobes and Bolhuis, 2007) and also prevent the restoration of normal adult song pitches following reinforcement-induced pitch changes (Canopoli et al., 2014).

Given its potential role in storing auditory memories, we investigated the role of NCM in the vocalizer identification task. We trained male and female zebra finches in an operant task that tested individual vocal recognition of several vocalizers using playbacks of songs and DCs. In previous work, we had shown that zebra finches excel at this task, forming long-lasting memories for vocal signature of new vocalizers very rapidly, requiring fewer than 10 exposures. Their memory capacity is large as they can remember up to 50 distinct vocal signatures simultaneously (Yu et al., 2020). We postulated that pairing NCM lesions with the vocalizer identification task might reveal nuances of the involvement of NCM in the acquisition and retrieval of auditory memories. We assessed whether bilateral neurotoxic lesions to NCM affected the ability to recall previously learned vocal signatures and to learn a new set of vocal signatures (i.e., from different birds). By analyzing task performance during the initial exposures to vocal signatures before and after lesion, we distinguished between the recall of previously learned vocal signatures and the learning or relearning of those vocal signatures. We also compared the effect of NCM lesions with lesions of the song nucleus HVC (used as a proper name). HVC has been shown in many studies to play an essential role in the generation of the complex motor pattern needed for song production (Hahnloser et al., 2002; Long and Fee, 2008). We found that lesions to NCM impair the ability of zebra finches to recall previously learned vocal signatures but not necessarily their ability to learn and discriminate between those calls. In contrast, lesions of HVC do not seem to have any effect on retrieval of auditory memories nor the learning of new vocal signatures in the vocalizer identification task.

Materials and Methods

Ethics statement

Animal experiments were approved by the Institutional Animal Care and Use Committee of the University of California, Berkeley (UC Berkeley; protocol number AUP-9157) following the guidelines of the National Institutes of Health and the Association for Assessment and Accreditation of Laboratory Animal Care International.

Animals

Twenty-one adult domesticated zebra finches, *Taeniopygia guttata* (13 male, 8 female), raised in our laboratory colony were used in these experiments. Before the experiments, the birds were housed in large aviaries where they interacted with many other adult and young conspecifics. The birds were divided into the following experimental groups: NCM (n = 10; 5 male, 5 female), HVC (n = 7; 5 male, 2 female), and Control (n = 4; three male, 1 female). In addition, we excluded a single male bird before performing the lesion procedure because he failed to learn the operant conditioning task in typical delays (~1 week). Three additional birds were trained but died during the surgical procedure.

Experimental design

In a baseline training phase, each bird was trained in an operant task to recognize the song and distance call of several conspecifics using the procedure described in our previous work (Yu et al., 2020). In brief, subjects were trained to trigger short (6 s) playbacks of distance call stimuli (DC) or song stimuli (Song) from vocalizers associated with rewarded (Re) and nonrewarded (NoRe) outcomes (Fig. 1*B*; see below, Song and distance call stimuli). Birds learn to terminate NoRe trials through key pecks, which increase reward rate by immediately skipping to the next trial (see below, Operant conditioning). Task performance was measured by comparing the tendency for a subject to interrupt NoRe playbacks to their tendency to interrupt Re playbacks. This quantity was represented by an odds ratio (OR), that is, the odds of interrupting a NoRe trial, $p(int|NoRe)/(1 - p(int|NoRe))$, divided by the odds of interrupting a Re trial, $p(int|Re)/(1 - p(int|Re))$. An OR significantly >1 indicates successful task performance, whereas a score of one indicates performance at chance level.

Training on a stimulus set (S1) during this prelesion learning phase took place over 5 d, with the size of the stimulus set increasing each day, up to 16 vocal signatures total, 8 Re and 8 NoRe (see Fig. 3*A*; Table 1). We refer to these learning procedures using an increasing number of vocal signatures as a "learning ladder." On day 1 (d1) the learning ladder begins with playbacks from songs or calls from one Re vocalizer and one NoRe vocalizer [one vs one (1v1)], then increases to playbacks of four Re and 4 NoRe (4v4) vocalizers on the second day. When the set size increases beyond 4v4, additional stimuli are introduced gradually, first with a day of 6v6-d1 for DC or 8v8-d1 for Song, during which novel stimuli are presented three times more frequently than previously learned stimuli, followed by 6v6-d2 for DC and 8v8-d2 for Song, during which all stimuli of one reward contingency were played at equal rates.

After bilateral lesions to NCM, HVC, or sham lesions in controls, birds were then retested on these previously experience stimulus sets [Set 1 (S1)] in a postlesion recall phase. Here, one session of 1v1 was used to verify that the birds were still capable of doing the task (Fig. 1, method; see Fig. 11, results) followed immediately by two sessions of the full 6v6-d2/8v8-d2. Finally, during the postlesion learning phase, subjects were trained to recognize a novel set of vocal signatures, Set 2 (S2), that they had not been exposed to before lesion. The postlesion learning of S2 followed the same schedule as the prelesion learning phase. Distance call sets and Song sets were not mixed but were tested in separate ladders within S1 and S2.

Operant conditioning

Detailed descriptions of the operant task and apparatus can be found in Elie and Theunissen (2018) and Yu et al. (2020). In brief, subjects were placed in an operant chamber set up with a speaker, food hopper, water bowl, and backlit pecking key (Med Associates). Subjects were tasked with discriminating between a set of Re and NoRe individuals based on the playback of their vocalizations. Subjects initiated trials by pecking on the backlit key, which triggers a 6 s stimulus playback (Fig. 1*B*). After 6 s, stimulus playback ends, and either nothing happens (NoRe trial), or a reward is given by raising the food hopper for 12 s (Re trial). Alternatively, a subject may peck the key at any time during the 6 s playback period to terminate the trial and begin a new trial with a random stimulus. In this case, no food reward will be given regardless of whether

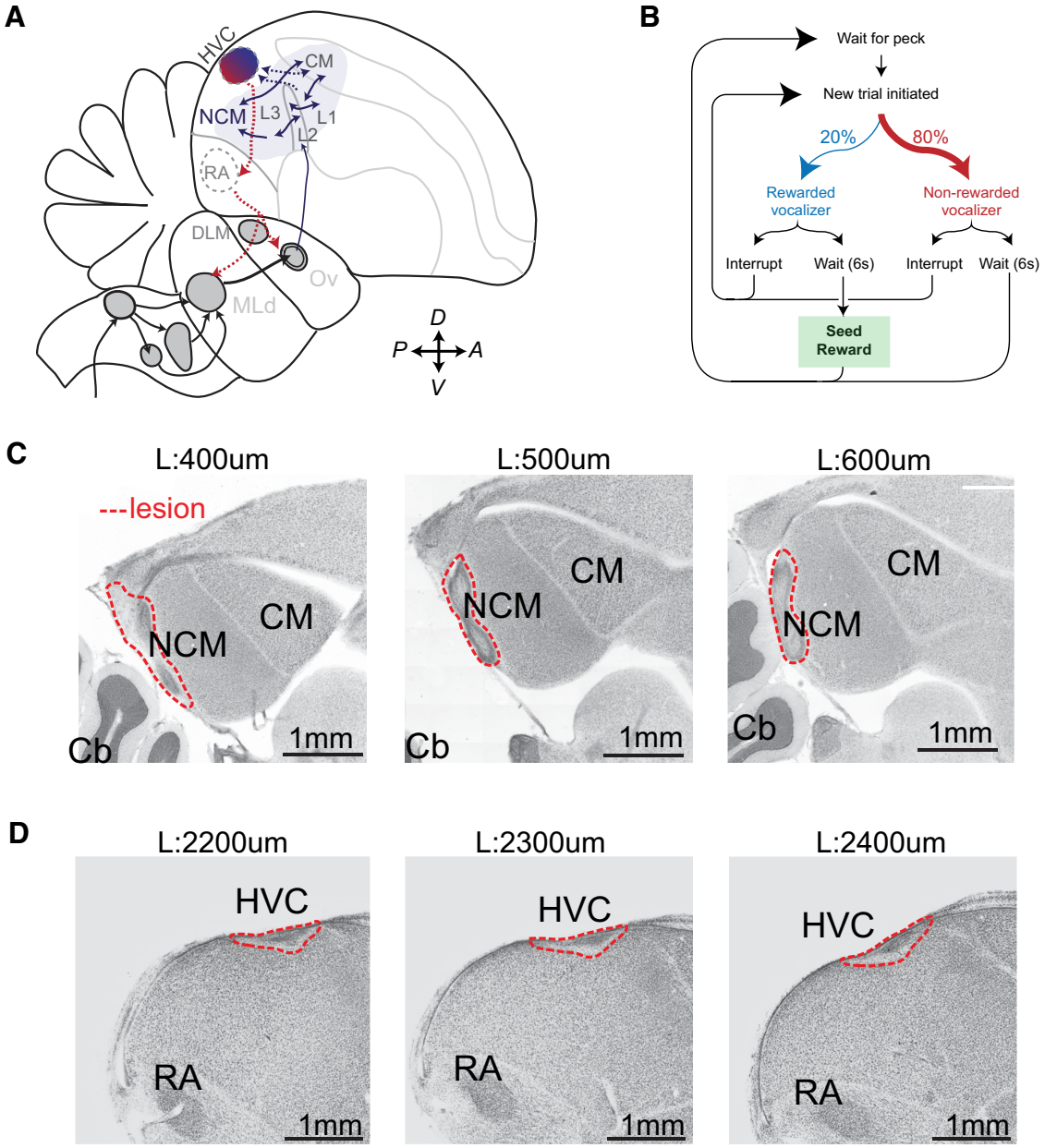

Figure 1. Zebra finch auditory and vocal motor pathways, task structure, and lesion images. ***A***, Diagram of the zebra finch auditory pathways in blue (solid) and vocal motor pathways in red (dashed). ***B***, Task diagram for behavioral conditioning. Subjects initiate trials and hear a 6 s stimulus playback of a rewarded or nonrewarded vocalizer. At the end of the playback, subjects will either receive a food reward or nothing. The subject can peck again during the playback (Interrupt) to terminate the trial and begin a new trial. Interrupting allows the bird to get rewards more rapidly if it mostly interrupts nonrewarded stimuli. Note that if a trial is interrupted, the bird does not gain information on whether the stimulus is rewarded or nonrewarded. ***C***, ***D***, Nissl-stained images of NCM and HVC lesions. Cb, Cerebellum; RA, robust nucleus of the arcopallium. Red dotted line indicates the approximate extent of the lesion. Details on injection coordinates, injection protocols, lesion sizes, and off-target lesions for all birds and all injections sites can be found in Extended Data Tables 1-1 and 1-4.

the initial trial was Re or NoRe. To maximize the rate at which reward is received in a session, subjects learn to skip stimuli that are recognized as NoRe to avoid the full waiting period and move on to the next trial. By design, 20% of trials are rewarded, whereas 80% of trials are not rewarded.

Between sessions, subjects are food restricted with access to water but limited seed to maintain motivation. Subjects were weighed before and after every test session, and seed consumed in a daily session was measured and supplemented at the end of day such that the birds maintained their weight within 10% of their weight at the start of the experiment. Daily handling of subjects did not seem to affect the motivation of the birds or ability to do the task once they became comfortable with the experiment chamber. Once trained, birds were able to get all their daily food allowance (2 × g/day) during the testing period.

The birds learn to use the apparatus during a shaping session that lasts ~1 week. During the shaping session, the bird first learns to associate pecking of the key with sounds from a Re vocalizer and food reward, followed by a gradual increase in presentations of a NoRe vocalizer. The initial shaping task involves the discrimination of two clearly distinct song stimuli. We have also performed control experiments, clearly showing that the apparatus is not providing any extraneous clues that the birds could use to distinguish Re from NoRe trials (Elie and Theunissen, 2018).

Song and distance call stimuli

Audio recordings of Song and DC stimuli originated from multiple labs and were described in Yu et al. (2020). Song vocalization recordings were from 32 male zebra finches from the Theunissen laboratory at UC

Table 1. Description of the learning ladder

Day	DCs	Songs	Description
1	1v1	1v1	1 Re vocalizer, 1 NoRe vocalizer
2	4v4	4v4	3 Re vocalizers added (4 total) and 3 NoRe vocalizers added (4 total)
3	6v6-d1	8v8-d1	DCs, 2 Re vocalizers added (6 total) and 2 NoRe vocalizers added (6 total). Songs, 4 Re vocalizers added (8 total) and 4 NoRe vocalizers added (8 total). Newly added vocalizers are played 4 times more frequently for their respective reward category.
4	6v6-d2	8v8-d2	No new vocalizers added. All Re vocalizers played at the same frequency. All NoRe vocalizers played at the same frequency.
5	6v6-d2	8v8-d2	Repeat of day 4.

Description of 1 week learning ladder training procedure in this article and illustrated in Figure 2*A*.

Berkeley, the Perkel laboratory at the University of Washington, and the Leblois laboratory at the Bordeaux Neurocampus in France. DC vocalizations came from 24 zebra finches (12 male and 12 female), all from our colony at UC Berkeley. Vocalizations used as stimuli were recorded as part of previous experiments in the laboratory, and the vocalizers were unknown to the subjects in the present study. The 12 male DCs were produced by a subset of the males also used in the song stimulus set; however, reward associations were randomized (seven had switched reward contingency, five the same). Previous work has shown that it is unlikely that zebra finches can generalize vocalizer identity from one call type to another (Elie and Theunissen, 2018); thus, for the purposes of this study the DC and Song stimuli recorded from the same males are treated as separate individual vocal signatures. However, it is possible that these songs may include some acoustic elements that resemble the DC.

For each individual vocalizer used as a stimulus in the task, we prepared 10 unique stimulus files composed of calls or song motifs from that individual. In this article, we use the term "vocalizer" to refer to the collection of 10 unique stimulus files of calls or song motifs from the same individual, as opposed to "rendition," which refers to a single stimulus file of the 10. These multiple renditions were used so that specific extraneous acoustic features of a particular stimulus file irrelevant to encoding vocalizer identity (e.g., length, intensity, and background noise) could not be used as a reward cue. In Yu et al. (2020), we showed that birds generalized their behavior over the 10 renditions from the same individual, and so our analyses of task performance in this study are done at the vocalizer level.

Song stimuli were constructed by combining three sampled motifs from one individual, whereas DC stimuli were constructed by combining six sampled DCs. Most introductory notes were removed from song motifs to avoid great variability in stimulus duration. Each DC sampled consisted of a single call or in some cases a pair of calls if the vocalizer did not normally produce single, isolated distance calls. These examples were arranged with pseudorandom intervals such that the duration of the file would be exactly 6 s long. Amplitudes of the audio files were then normalized within stimuli of the same type, that is, Songs or DCs, and played such that the average sound pressure level was ~70 dB as measured by a handheld dB meter in the center of the operant chamber. Example spectrograms of stimulus playback files can be found in our previous work (Yu et al., 2020).

Prelesion and postlesion training and testing

The full stimulus sets in these experiments included playbacks of the songs of 16 different vocalizers or the distance calls of 12 different vocalizers. Our learning ladder procedure, described above (and summarized in Table 1), is designed to gradually introduce playbacks from more vocalizers to a subject each day and thus quantify learning rates and memory capacity. This learning ladder was performed before lesion or sham surgeries for songs and DC separately (Song S1 and DC S1) and postlesion on a novel set of vocal signatures for songs and DC separately (Song S2 and DC S2). In addition, postlesion birds were retested soon after lesion on the recall of the S1 vocal signatures learned in the prelesion training. Stimulus sets are summarized in Table 2.

Table 2. Description of the stimulus sets

Set name	# Vocalizers	Renditions/vocalizer	Description
Song S1	16	10	Songs first learned before lesion, retested after lesion
DC S1	12	10	DCs first learned before lesion, retested after lesion
Song S2	16	10	Songs first learned after lesion
DC S2	12	10	DCs first learned after lesion

Description of the stimulus sets used in the task.

Lesions

Surgery. Following the behavioral tests of S1, subjects received either bilateral NCM lesions (n = 10, 5 males, 5 females), bilateral HVC lesions (n = 7; 5 males, 2 females), or sham lesions (n = 4; 3 males, 1 female). Birds were food deprived for 1 h before anesthesia and orally administered 0.5 mg/kg meloxicam as analgesic. For the anesthesia, birds were induced at 2–4% isoflurane and maintained at ~1% isoflurane. For the surgery, the anesthetized birds lay on a heated sling with their head fixed using a stereotaxic apparatus (Kopf Instruments). A subcutaneous injection of lidocaine was administered, and head feathers removed. An incision along the midline was made and the skin retracted. A craniotomy was performed over each hemisphere around the desired injection coordinates using a dental drill. The midsagittal sinus was identified as the stereotaxic zero coordinate. A glass micropipette with tip diameter 25–40 μm was lowered using a hydraulic micromanipulator. Bilateral excitotoxic lesions were made using 2% *N*-methyl-DL-aspartic acid (NMA) solution (six NCM birds) or 0.7% ibotenic acid (four NCM birds, seven HVC birds) dissolved in PBS. Stereotaxic coordinates for NCM and HVC were taken from the existing literature and adjusted based on our own histologic verification of lesioning outcomes. Each injection at one coordinate (medial/lateral, rostral/caudal) was performed at either one or two depths. The pipette was first lowered to the deepest injection site, and solution was released in many small puffs (4–12 nl) using the Nanoject II system (Drummond Scientific). The pipette was left in place for 5 min after each site before retraction to the next injection site or out of the brain. The stereotaxic coordinates used for each bird are fully described in Extended Data Tables 1-1 and 1-2. The method of sham lesion varied in the Control group. In two Control subjects, dye was injected using the same coordinates used for the NCM group. A third Control was originally in the NCM group, but the lesion was found to be exceptionally small, most likely because of a clogged pipette during surgery. A fourth Control was originally in the HVC group and targeted with ibotenic acid, but the lesion was found to be off target in both hemispheres. At the conclusion of surgery, craniotomies were covered with Kwik-Cast, the surface sealed with Vetbond, and Bacitracin ophthalmic ointment was applied to prevent infection.

After the operation, subjects were returned to their home cages with food provided *ad libitum*. Recovery time varied by subject and ranged from 2 full days to 7 full days (mean = 4.8 d). After recovery, subjects continued with the operant task starting with the re test of stimulus sets S1 as described above (see Fig. 3*A*).

Lesion size quantification. Lesion sizes were determined by a researcher who was naive to subject identity. Brains were fixed using 4% PFA transcardial perfusions, followed by sucrose cryoprotection and sliced at 50 μm on a freezing microtome. Alternating sections were mounted, Nissl stained (Cresyl Violet), and scanned at the Cancer Research Lab (CRL) Molecular Imaging Center at UC Berkeley. Calibration of slice dimensions was automatically performed by the Zeiss acquisition software, Zen Blue. The same software was used to hand draw lesion areas on each slice. Linear interpolation was performed to estimate the lesion area between slices, and lesion volume was then numerically integrated across all slices. Eugen et al. (2020) performed a phylogenetic analysis of NCM and argued that dopaminergic innervation suggests NCM extends to ~1.1 mm laterally in the zebra finch, which is also the boundary the Zebra Finch Expression Brain Atlas (zebrafinchatlas.org) uses. NCM was

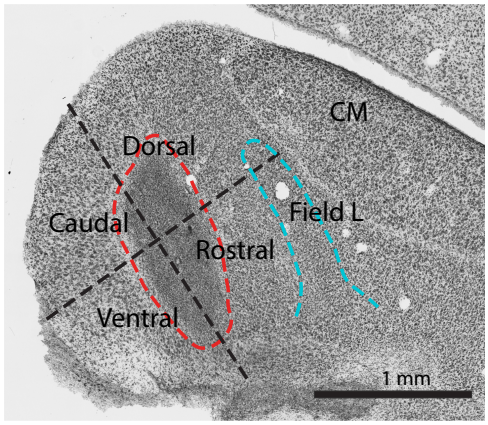

Figure 2. Division of NCM into four quadrants. The NCM was divided into four quadrants using the midpoints of the dorsal/ventral and caudal/rostral extent of the medial-caudal nidopallium as illustrated on this parasagittal slice. The caudal/dorsal to rostro/ventral axis of this partition was drawn to be parallel to the major axis of the field L region, shown here with blue dashed lines. The red dashed lines delineate the NCM lesion observed in this bird. The slice shown here is 0.8 mm from the midline. CM, Caudal mesopallium.

defined as the nidopallium from the midline to 1100 μm lateral and excluding Field L. Multiple landmarks were used to estimate distance from midline when slices close to the midline were damaged or ambiguous, including the dorsal medial arcopallium, HVC, the shape of the lobe containing NCM and Field L, and the laminae between nidopallium and mesopallium. Because NCM is a large region with potentially distinct anatomic, histologic, and potentially functional divisions, we also subdivided into four quadrants as shown in Fig. 2. The volume of the lesion within each quadrant was also estimated.

In several subjects, areas beyond NCM were partially lesioned. For most subjects, this off-target lesioning was minimal (Extended Data Table 1-4). Off-target lesions in caudal mesopallium (CM), Field L, hippocampus, cerebellum, and nidopallium caudolaterale (NCL) were labeled by a blinded researcher on a four-point scale, where zero indicated no lesioning, and three indicated potentially extensive lesioning.

In all but one HVC lesioned bird the HVC lesions clearly encompassed the vast majority of both HVCs and resulted in either no or substantially degraded song. In one male subject, lesions missed HVC completely, and the song of the subject was not degraded (see Fig. 12); this subject was reassigned to the Control group for analyses.

Learning curves and statistical analyses

Uninterrupted trials. We used uninterrupted trials to measure how the task performance of a subject changed as a function of experience with the reward contingency of a given vocalizer. Because of the asymmetric treatment of interrupts and noninterrupts of our task structure (Fig. 1*B*), birds could only learn whether a stimulus was rewarded when they refrained from interrupting the playback; interrupting a trial always triggers the next trial without providing a food reward, regardless of the reward contingency of the current stimulus. In contrast, uninterrupted trials conclude with either food reward or nothing. Note that by simple exposure, all trials could provide indirect information that could be used for the formation of perceptual categories without any reinforcement. Nonetheless, in this task the information obtained during uninterrupted trials is the only one that could be used in reinforcement learning. We will therefore use the number of uninterrupted trials, k, to quantify the potential amount of information a subject can gain about stimulus–reward contingencies from a given reference point (where $k = 0$ is chosen). In other words, analysis using k as the independent variable lets us describe how additional examples of reward/nonreward influence task behavior. Low values of k describe the regime in which birds have not been exposed to many examples and must rely on prior experience; as k increases, the birds have had more opportunities to learn or relearn reward contingencies and can rely on both past and current information.

A separate count is kept for each (subject, vocalizer) pair, which starts at $k = 0$ and increments independently for each vocalizer each time that the subject does not interrupt a trial associated with that vocalizer. This count of k can extend over multiple sessions of our ladder structure if playbacks from the same vocalizers are tested over multiple days (Fig. 3). Thus, uninterrupted trials allow us to make a fair comparison of performance between playbacks from a vocalizer introduced on day 1 to playbacks from vocalizers introduced on days 2 and 3. To compare learning during different time periods we can choose where to set $k = 0$. For example, to compare performance before and after lesion, we took scores on the second-to-last day before lesion and compared them with scores taken with k set to 0 on the first session after lesion.

In this study, we analyze performance after lesion at two scales, early (for all $k \leq 10$), and late (sessions 6v6-d2 and 8v8-d2, up to the maximum value of k obtained for all subjects). For the early period, we estimate performance curves of OR as a function of k to examine the effect of experimental conditions during this period. Note that some vocal signatures are first presented on different days (e.g., after lesion, only two vocal signatures are presented in the initial 1v1 session, whereas the remaining vocal signatures are presented in the subsequent 6v6-d2/8v8-d2 sessions, as shown in Figure 3, *B* and *C*. For this reason, and when assessing learning performance, uninterrupted trials with a given value of k may fall on different sessions for different vocal signatures. However, when we assessed recall (or relearning) in late session, we used the start of the day 2 session to begin counting the uninterrupted trials for all vocal signatures.

Estimation of learning curves and their statistical modeling. We estimated group average learning curves using the following procedure. First, we gathered trial data from a set of subjects S (e.g., subjects with NCM lesions) responding to a stimulus set of vocal signatures $V \in \{V_{\mathrm{Re}}, V_{\mathrm{NoRe}}\}$, where V_{Re} is the set of rewarded vocal signatures, and V_{NoRe} is the set of nonrewarded vocal signatures. For each uninterrupted trial bin k, we defined the number of trials in that bin to be the integer T_k^{sv}. To be precise, T_k^{sv} is the empirical number of trials between the kth (exclusive) and $(k+1)$th (inclusive) uninterrupted trial of a vocalizer v by subject s (illustrated in Table 3). This bin typically consists of $T_k^{sv} - 1$ interrupted trial(s) and 1 uninterrupted trial.

To get the overall counts of interrupted trials of a subject for either Re or NoRe vocalizer, we sum the counts over all vocal signatures with the same reward contingency as follows:

$$n(\mathrm{int}|s, V, k) = \sum_{v}^{V} n(\mathrm{int}|s, v, k) = \sum_{v}^{V} (T_k^{sv} - 1).$$

Here, v indexes the vocalizers belonging to either $V \in \{V_{\mathrm{Re}}, V_{\mathrm{NoRe}}\}$. Similarly, the number of uninterrupted trials is simply as follows:

$$n(\mathrm{unint}|s, V, k) = \sum_{v}^{V} 1.$$

At this stage, probabilities of interruptions and OR can be obtained for a given subject and k as follows:

$$p(int|s, V, k) = \frac{n(\mathrm{int}|s, V, k)}{n(\mathrm{int}|s, V, k) + n(\mathrm{unint}|s, V, k)}$$

$$OR(s, k) = \frac{n(\mathrm{int}|s, V_{\mathrm{NoRe}}, k) * n(\mathrm{unint}|s, V_{\mathrm{Re}}, k)}{n(\mathrm{unint}|s, V_{\mathrm{NoRe}}, k) * n(\mathrm{int}|s, V_{\mathrm{Re}}, k)},$$

and Fisher's exact test can be performed to assess whether the OR is significantly >1 for a given subject and k.

The group average OR is obtained by further summing the counts across subjects as follows:

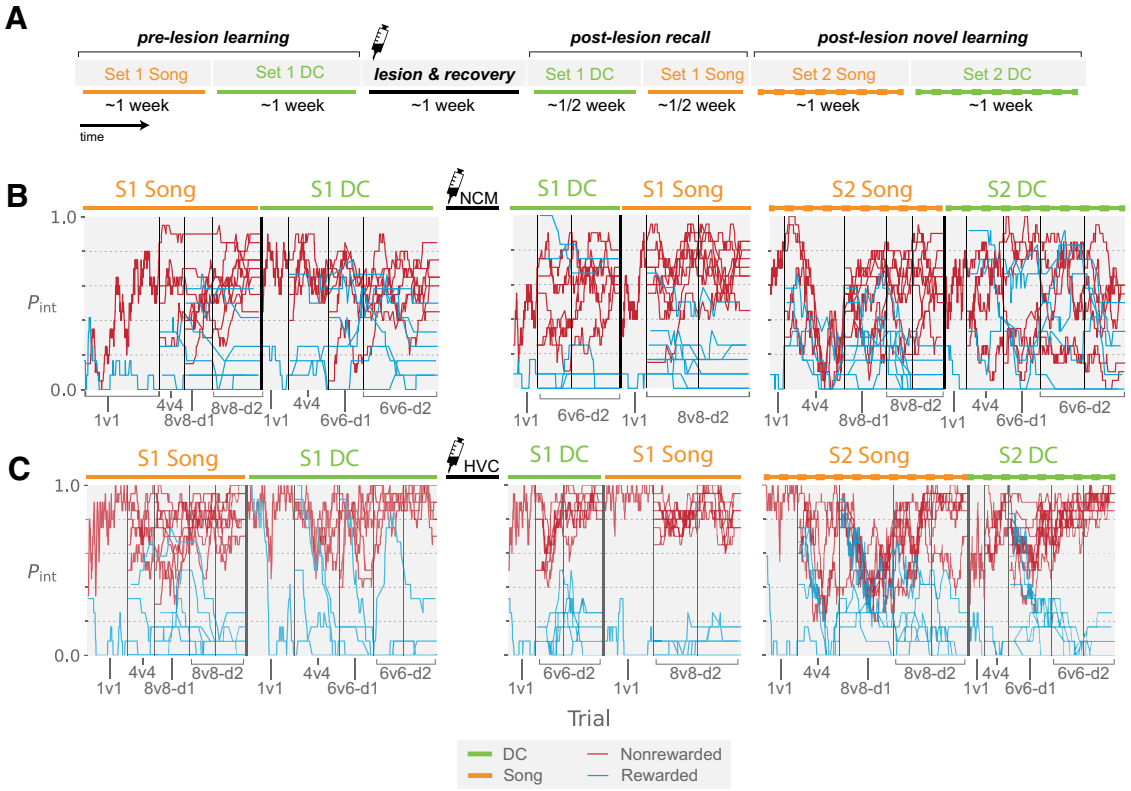

Figure 3. Learning ladder prelesion and postlesion. ***A***, Schematic of the three phases of the experiment in an example subject. First, the subject is trained on a set of songs followed by a set of DCs (prelesion learning of S1). It then undergoes surgery during which NCM or HVC is lesioned (or saline/dye injected in controls) and up to 1 week of recovery. Next, it is tested on recall of the previously learned DCs and songs in succession (postlesion recall of S1). Finally, it is trained and tested on a novel set of songs and DCs in succession (postlesion learning of S2). The alternating pattern between Song and DC sets was swapped in half of the subjects. ***B***, ***C***, Example traces showing the probability of interruption per vocalizer during the three phases of the experiment in two subjects. In ***B***, a subject with a focal NCM lesion shown in Fig. 1 and in ***C***, a subject with an HVC lesion. Each continuous line shows the probability of the subject interrupting the playbacks from the same vocalizer, calculated in a sliding window (blue, rewarded vocalizer, 12 trial bin width; red, nonrewarded vocalizer, 20 trial bin width). Each day is separated by a thin vertical line. New lines appearing in the middle of the weeks correspond to the addition of playbacks from an additional vocalizer belonging to a particular set. The labels and the color and shape of the horizontal lines above the plots indicate the call type being tested and whether it is S1 and S2 (Songs, orange; DCs, green; S2 with bumpy line). Annotations below the plot show the number of vocal signatures tested each day.

$$n(int|V,k) = \sum_{s}^{S} n(\text{int}|s,V,k)$$

to obtain the following:

$$p(int|V,k) = \frac{n(\text{int}|V,k)}{n(\text{int}|V,k) + n(\text{unint}|V,k)}$$

$$Log_2(OR(k)) = Log_2\left[\frac{n(\text{int}|V_{\text{NoRe}},k) * n(\text{unint}|V_{\text{Re}},k)}{n(\text{unint}|V_{\text{NoRe}},k) * n(\text{int}|V_{\text{Re}},k)}\right].$$

The SEs of the OR(k) is given by the following:

$$SE(OR(k)) = \sqrt{\frac{1}{n(\text{int}|V_{\text{NoRe}},k)} + \frac{1}{n(\text{int}|V_{\text{Re}},k)} + \frac{1}{n(\text{unint}|V_{\text{Re}},k)} + \frac{1}{n(\text{unint}|V_{\text{NoRe}},k)}}.$$

These are the group average Log_2OR that are shown in the learning curves on Figures 4, 5, 6, 8, 9. Note that this is a weighted average that gives more weights to subjects that have a higher number of interrupted trials (for either reward contingency) and thus for which the estimates of the probabilities of interruption versus noninterruption are more accurate.

To perform statistical analyses that explicitly address the independent assessment of these probabilities given the data of each subject, estimates of the log(Odds) were obtained using mixed-effect generalized linear models for binomial distributions (also known as mixed-effect logistic regression). In the simplest case when one is fitting a learning curve for a single condition (e.g., the Control lesioned birds on S1 before lesion), the mixed-effect model would be written in R compact equation notation as follows:

$$(n_{\text{int}}, n_{\text{unint}}) \sim k * V + ((1 + k * V)|Subject).$$

Here, V is a factor that takes on two levels, Re and NoRe (or True and False). The second term in the sum on the right side of the equation specifies the random effects. The mathematical formulation for this statistical model is given by the following:

$$\log\left(\frac{p(\text{int}|s,V,k)}{1-p(\text{int}|s,V,k)}\right) = b_0 + b_1 k + b_2 V + b_3(k * V) + (\alpha_{0,s} + \alpha_{1,s}k + \alpha_{2,s}V + \alpha_{3,s}k * V) + \varepsilon,$$

where b_i are the fixed-effect coefficients, $\alpha_{i,s}$ are the random coefficients, and ε is the noise. V takes on the value of zero for the NoRe stimuli and one for the Re stimuli. The coefficients of these models are fitted by maximum likelihood assuming centered and normal distributions of random effect coefficients. The fixed-effects coefficients provide fits of the group average log(Odds) and log(OR) as a function of k and reward contingency. More explicitly, the log(Odds) for the unrewarded stimuli across all subjects as a function of k is fitted with the following line:

Table 3. Illustration of calculating the number of interruptions as a function of uninterrupted trial

Trials	T_0	$p(int\|k=0)$	T_1	$p(int\|k=1)$	T_2	$p(int\|k=2)$
000	1	0	1	0	1	0
101101110	2	1/2	3	2/3	4	3/4
101101111X	3	1/2	3	2/3	9	8/9

Examples of uninterrupted trials counting for a single subject and vocalizer. Trials column, Example sequence of responses in the task; 1 indicates an interruption; 0 indicates a noninterruption, that is, uninterrupted trial; and X indicates the end of the dataset. T_k is the number of trials between the kth uninterrupted trial (exclusive) and the $(k+1)$th uninterrupted trial (inclusive). The columns $p(int|k)$ show the estimated probability of interruption for that bin, using the procedure described above, Materials and Methods, Learning curves.

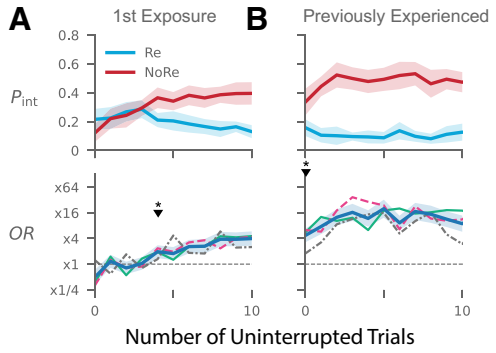

Figure 4. Robust recognition of learned vocal signatures before lesion. ***A***, Top, Lines show the average probability of interruption as a function of uninterrupted trials seen, relative to initial exposure to vocal signatures in the S1 stimulus set. Probability of interruption (P_{int}) to NoRe vocal signatures is shown in red, and Re vocal signatures in blue. Data are averaged over all subjects and vocalizers. Shaded regions show an estimate of 2 SEM by jack-knifing over subjects. Bottom, The difference between NoRe and Re interruption rates are summarized as an OR. OR $>$ 1 indicates successful task performance. Top, The first uninterrupted trial bin where OR is significantly greater than one is marked with a black triangle and asterisk. Averages of each experimental group are overlaid without shaded bars, and the group average is shown in solid blue with 2 SEM shaded. ***B***, Same as in ***A*** but counting uninterrupted trials from day 4 of the learning ladder (i.e., 8v8-d2 or 6v6-d2), after which each subject would have had at least one session of prior exposure to each vocalizer in S1.

$$\log\left(\frac{p(\text{int}|\text{NoRe},k)}{1-p(\text{int}|\text{NoRe},k)}\right) = b_0 + b_1 k.$$

Similarly, the log(Odds) for the rewarded stimuli across all subjects as a function of k is fitted with the following line:

$$\log\left(\frac{p(\text{int}|\text{Re},k)}{1-p(\text{int}|\text{Re},k)}\right) = b_0 + b_1 k + b_2 + b_3 k.$$

Thus, the log(OR) as a function of k is given by the difference in the log of Odds, or as follows:

$$\log OR(k) = b_2 + b_3 k.$$

Here, the coefficients b_2 and b_3 are the estimates of the intercept and slope of lines fits to the learning curves plotted in Figures 4, 5, 6, 8, 9 and are shown explicitly in Table 4. The Wald test on these coefficients using the SE obtained from the mixed-effect statistical modeling estimates can then be applied to test whether the intercept or slope are different from zero.

More generally, we are interested in comparing the fits of log(OR) across treatment conditions such as differences in intercepts and slopes for NCM versus Control lesions birds. For this purpose, we compare the prediction performance of nested mixed-effect models that include or exclude Treatment as an additional factor. In R compact equation notation, we would compare the reduced model corresponding to a single line fit for the log(OR) and given by the following:

$$(n_{\text{int}}, n_{\text{unint}}) \sim k * V + ((1 + k * V)|Subject)$$

to a full model corresponding to a different line being fitted for each Treatment condition as follows:

$$(n_{\text{int}}, n_{\text{unint}}) \sim k * V * Treatment + ((1 + k * V)|Subject).$$

A likelihood ratio test can be performed to assess whether the fit by the full model is superior to the fit given by the reduced model. If so, the Wald test on specific coefficients provides estimates of which intercepts and slopes are significantly different from each other for different conditions (e.g., Control vs NCM lesions). Specifying the mathematical formulation for the full model is essential for the correct identification of the relevant coefficient for performing such statistical analyses. This mathematical formulation is not given here, but it is an extension of what has been shown above, and it can also be found in the R scripts provided (see below, Data availability). Statistical analyses and estimates of the effect size were also performed by grouping HVC and Controls (n = 11) to provide a better sample size match to the NCM group (n = 10).

Estimating $log_2(OR)$ for a range of uninterrupted trials

We also evaluate average task performance during particular time windows such as for $k \leq 10$ to assess performance during this initial learning (S1 before lesion) or recall/relearning (S1 after lesion) period or such as in the last days to testing (days 2). For these analyses, we simply sum the count of interrupted trials and uninterrupted trials up to a specified maximum value of uninterrupted trials as follows:

$$n_s(\text{unint}|v, k \leq k_{max}) = k_{max}$$

$$n_s(\text{int}|v, k \leq k_{max}) = \sum_{k=0}^{k\_max} (T_k^{sv} - 1).$$

As above, Fisher's exact tests can then be used to assess whether ORs are significantly different from zero for particular subjects, and mixed-effect generalized linear models can be used to test for particular effects. For example, the statistical effect of treatment will be assessed by comparing the likelihood of the following two nested models:

$$(n_{\text{int}}, n_{\text{unint}}) \sim V + ((1 + V)|Subject)$$

$$(n_{\text{int}}, n_{\text{unint}}) \sim V * Treatment + ((1 + V)|Subject).$$

In the results, we first used the interruption counts for all stimuli (and all vocalizers) regardless of which vocalization type (Song and DC) was used. To assess differences in the way birds treated these two types of vocalizations, and whether lesions had different effects on recall/recognition of each, we then also performed the analyses separately for songs and DCs. Baseline interruption rates for each vocalization type were measured per subject as an average across vocalizers for that type, and group averages were computed over subjects. To test whether vocalization type added additional explanatory power to performance scores, and whether lesions to NCM or HVC affected Song/DC differently, we applied linear mixed-effects statistical modeling by comparing base models with alternate models with fixed intercepts for Song/DC and interaction terms between Song/DC and the lesion group.

Task performance and lesion volume

The relationship between task performance and lesion volume was modeled with linear bivariate regression, with lesion size as a regressor and decreases in task performance caused by the lesion as the predicted variable. The decrease in task performance was the difference in the log(OR) obtained for the S1 stimulus set between after and before lesion estimated on the final day (d2). The log(OR) for each subject were estimated using Equation 1. The p values were corrected for multiple comparisons

using Bonferonni for the quadrant analysis, which involved four separate bivariate regressions.

Code and software

The operant conditioning box and system is operated using a custom fork of the Python-based PyOperant software (https://github.com/theunissenlab/pyoperant), originally developed by Justin Kiggins and Marvin Thielk in Timothy Gentner's laboratory at the University of California, San Diego (https://github.com/gentnerlab/pyoperant).

Data availability

Code for analysis was written in Python and R. Code, documentation, data files, Python dependencies, and Jupyter notebooks used for analysis and production of figures in this article can be found online at https://github.com/theunissenlab/zebra-finch-memory-lesions. Statistical tests, for example, Fisher's exact test and bivariate regression, were run using the Python packages scipy and statsmodels and the generalized mixed-effect model using the R library lme4.

Results

To assess whether lesions to NCM and HVC affect auditory memory functions needed for identifying a vocalizer, we tested zebra finches in an individual vocal signature recognition task using songs and DCs from different birds. To measure the effect of the lesion on memories, we first compared the discrimination performance of the bird after lesion to its performance on the same set of stimuli learned before lesion. This initial stimulus set is referred to as S1. To measure the effect of the lesion on learning, we then tested the birds on a new set of playbacks from different vocalizers (called S2) that they had not been exposed to before lesion. Birds were treated with bilateral lesions to either NCM or HVC, and sham lesions were performed in Controls. To test for recognition, we used a modified go/no-go operant task in which subjects optimize their reward rate and simultaneously demonstrate successful recognition by interrupting NoRe vocal signatures more frequently than Re vocal signatures (see above, Materials and Methods). We found that although vocalization type (i.e., Song or DC) have different rates of baseline interruption, the learning curves and effects of lesions were similar (see below, Songs versus distance calls). For this reason, the initial section and figures in the article describe the results for Song and DC combined and results for each call type (see Figs. 8-11).

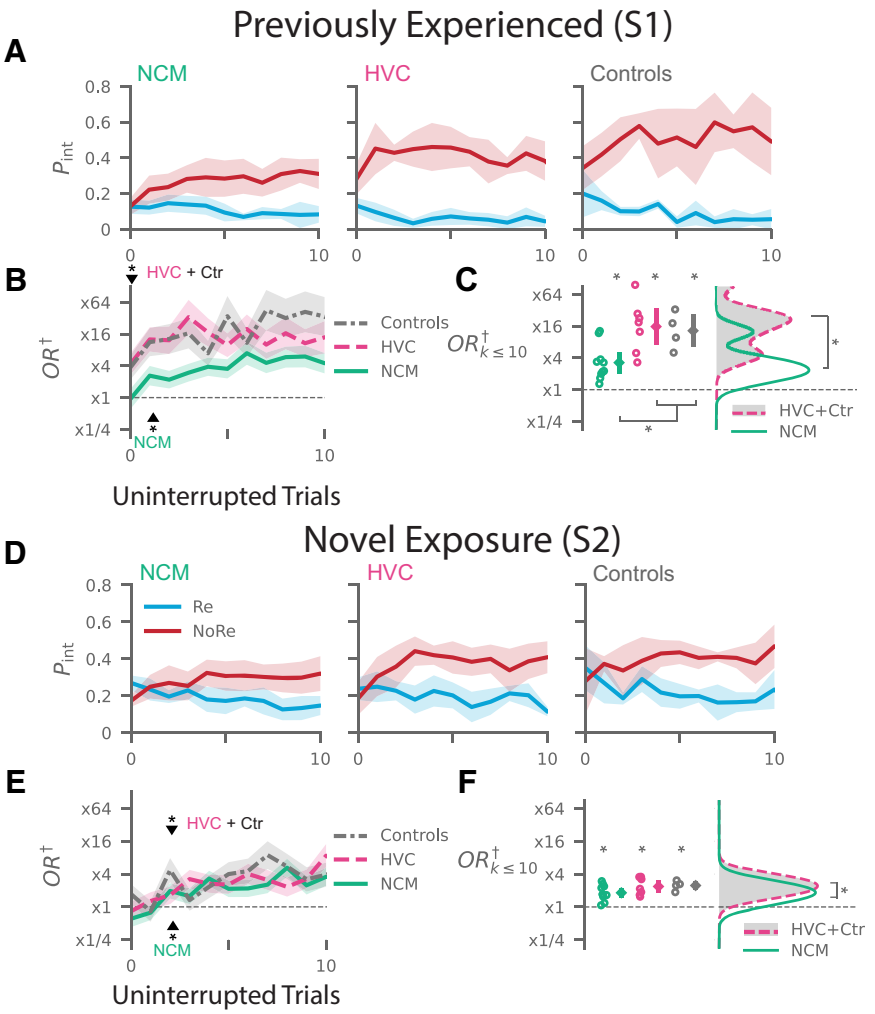

Figure 5. Effect of lesions on the recognition of previously experienced (S1) and novel stimuli (S2). The effect of lesion on discrimination performance for S1, (***A-C***), and for playbacks from novel vocalizers, Novel Exposure (S2), (***B-F***). ***A***, ***D***, Average probability of interruption as a function of uninterrupted trials after lesion for vocalization stimuli originally learned before lesion. Interruption rate to NoRe vocal signatures is shown in red, and Re vocal signatures in blue. Data are averaged over all subjects and vocalizers. Shaded regions show an estimate of 2 SEM by jack-knifing over subjects. These curves can also be compared with the preoperation data shown in Figure 4*B*. ***B***, ***E***, As in Figure 4 (bottom), the difference between NoRe and Re interruption rates in the first uninterrupted trials after lesion are summarized as an OR for each group. Shaded region indicates 2 SEM. The first uninterrupted trial where the OR is significantly different from one from there on is marked with a black triangle and an asterisk. HVC lesions and Controls are grouped in this analysis as they are statistically indistinguishable. ***C***, ***F***, The ratio of OR measured before the 10th uninterrupted trial after lesion (denoted as $OR^{\dagger}_{k \leq 10}$). Each scatter point shows the performance of one subject on all vocal signatures, with mean and 2 SEM to the right of each scatter. Significance markers shown above the data indicate $p < 0.05$ on one-sample t tests for $\log_2(OR^{\dagger}_{k \leq 10}) > 0$. Significance markers below plot shows result of two-tailed t test comparing the mean $\log_2$OR between the NCM group and the combined HVC + Control group. Kernel density estimates of these distributions are shown to the right.

Fast and persistent recognition of vocal signatures

We first characterized typical learning by examining average learning curves for both initial and late exposures to S1 during the prelesion phase of the experiment. Qualitative inspection of interruption rates per vocalizer over sessions and days (Fig. 3*B*, *C*) shows that interruption rates for Re and NoRe vocal signatures separate quickly, often within a single session, and are stable by the end of our learning ladder on days 4 and 5 corresponding to the 6v6-d2/8v8-d2 period (see above, Materials and Methods). To quantify the learning rate, we examined performance as a function of the uninterrupted trials as the birds can only gain feedback on the reward contingency when they don't interrupt (see above, Materials and Methods). During initial learning, birds took approximately four uninterrupted trials on average to distinguish NoRe from Re vocal signatures (Fig. 4*A*). Once the vocal signatures were learned, they could be recognized immediately in subsequent sessions (Fig. 4*B*); on the fourth day (the first session after which subjects had had exposure to all

Table 4. Intercepts and slope (*k*) coefficients of the line fitted to the log(OR) versus uninterrupted trials (*k*)

HVC lesioned + Control				NCM lesioned			
	Treat.	Coef.	SE		Treat.	Coef.	SE
Intercept	Pre-S1-d1	−0.55[a]	0.25	Intercept	Pre-S1-d1	−0.73[a]	0.32
k		0.28[a]	0.04	k		0.32[a]	0.05
Intercept	Pre-S1-d2	3.65[b]	0.55	Intercept	Pre-S1-d2	2.90[b]	0.56
k		0.04[b]	0.05	k		0.15[b]	0.06
Intercept	*Post-S1*	2.93[b]	0.35	Intercept	*Post-S1*	0.73[a]	0.33
k		0.25[b]	0.05	k		0.21[a]	0.05
Intercept	*Post-S2*	0.43[a]	0.15	Intercept	*Post-S2*	−0.05[a]	0.20
k		0.19[a]	0.03	k		0.20[a]	0.03

Mixed-effect linear models were used to obtain the coefficients and their SEs for the line that best fitted the relationship between the log(OR) and the number of uninterrupted trials. The font for the Treatment column is regular for the prelesion data (fitting the learning curves shown in Fig. 3) and italicized for the postlesion data (fitting the curves shown in Figs. 4, 5). The values of the coefficients and their SEs are superscripted by a and b; a corresponds to learning curves observed in initial exposure where the intercept is close to zero and the learning rate (slope) is higher, and b corresponds to learning curves observed for learned vocal signatures corresponds to learning curves observed for learned vocal signatures where the intercept is significantly above zero and the learning rate is lower. Coef, Coefficient.

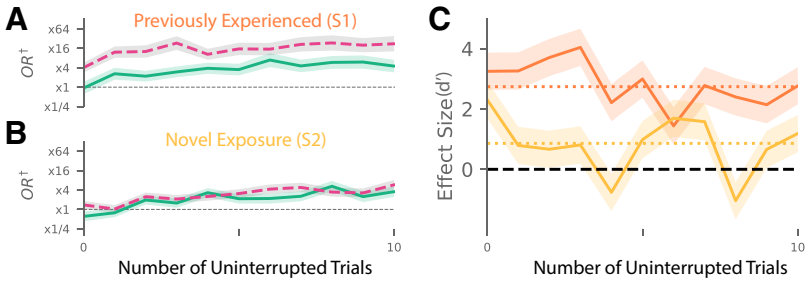

Figure 6. Effect size of NCM Lesions versus Control and HVC lesions for the recognition of Previously Experienced (S1) and Novel Stimuli (S2). ***A***, ***B***, The odds ratio quantifying the difference in interruption rates for rewarded versus unrewarded stimuli is shown as a function of uninterrupted trials for NCM lesioned birds (green) and HVC lesioned birds combined with the Controls (HVC + CTRL; pink and gray). In ***A*** the learning/performance curves are shown for the previously experienced vocal signatures and in ***B*** for novel vocal signatures. These data are the same as in Figure 5, *B* and *E*, but after combining the data from HVC and Control birds. ***C***, The differences between the learning curves shown in ***A*** and ***B*** in units of SD (Effect size *d'*) are plotted also as a function of uninterrupted trials. Shaded ribbon corresponds to 2 SEM. The effect size for previously experienced vocal signatures is much larger than that for novel vocal signatures.

vocal signatures in S1 for at least 1 d previously), we confirmed that NoRe trials were more likely to be interrupted than Re trials before experiencing even a single uninterrupted trial [Fisher's exact test on interruption counts across all vocalizers and subjects, log2(OR) = 2.72, $p < 0.001$; Fisher's exact test per subject, 11/21 subjects with $p < 0.05$]. This result demonstrates that recognition memory for vocalizer ID persists across sessions.

NCM lesions affect discrimination performance of previously learned vocal signatures but do not prevent relearning

We next compared the prelesion performance described above with the learning/performance curves obtained shortly after lesions ($k \leq 10$) for the previously experienced stimulus set, S1 (Fig. 5*A–C*) and for a novel exposure to a set of playbacks from different vocalizers, S2 (Fig. 5*D–F*). Visual inspection of the interruption rates (Fig. 5*A*,*D*) and of the OR estimated from these interruption rates paired per bird (Fig. 5*B*,*E*) show little or no differences between the HVC lesioned and Control birds, but a remarked decrease in performance for the NCM birds for the previously learned stimuli only. To quantify these effects, we used generalized mixed-effects statistical models (mixed-effect logistic regression) with the number of interruptions and the total number of trials as the response variable, the reward contingency, the number of uninterrupted trials (k), the lesion type (Treatment) and their interactions as fixed effects and random intercepts and coefficients of the reward contingency and k for each subject. To test whether Treatment had a significant effect on task performance during this early remembering/learning period, we compared nested models that included versus excluded Treatment as a regressor (see above, Materials and Methods). For the previous learned stimuli (Fig. 5*B*), we found that the full statistical model that included Treatment was a better fit to the data than the model that excluded Treatment [three lines vs 1 line for fitting data; likelihood ratio test, $\chi^2(8) = 23.794$, $p = 0.0024$], showing an effect of Treatment.

Next, we performed *post hoc* tests for the intercept coefficients of the learning curves. We found that the intercept of the NCM lesioned subjects was significantly smaller than both HVC lesioned subjects [difference in $\log_2$(OR) = 2.69 ± 0.661 (SE); Wald test, $Z = 4.073$, $p < 0.001$)] and Control subjects [difference in log2 (OR) = 1.69 ±0.637 (SE); Wald test, $Z = 2.648$, $p = 0.0081$]. Thus, in the early trials after lesions ($k \leq 10$) NCM lesioned subjects performed on average worse than HVC and Control subjects. In contrast, the intercepts of Control and HVC subjects were not distinguished [difference in $\log_2$(OR) = −0.88 +-0.734 (SE); Wald test, $Z = -1.198$, $p = 0.2308$].

We repeated this analysis by combining the HVC and Control group ($n = 11$) to better match the sample size of the NCM lesioned group ($n = 10$) and increase our statistical power. Here again, the full statistical model that included Treatment (2 regression lines for NCM vs HVC + Control to fit Fig. 6*A*) was a better fit to the data than the single regression line model that excluded Treatment [likelihood ratio test, $\chi^2(4) = 15.478$, $p = 0.0038$]. The *post hoc* tests also revealed that the intercepts for the NCM versus HVC + Control lines were significantly different [differences in $\log_2$(OR) = −2.20 ± 0.499 (SE); Wald test, Z = −4.402, $p < 0.001$].

In contrast, for playbacks from novel vocalizers (Fig. 5*D–F*), the full statistical model that included Treatment was not statistically distinguishable from the model that excluded Treatment [three lines vs one line for fitting; Fig. 5*E*; likelihood ratio test, $\chi^2(8) = 14.014$, $p = 0.0814$]. This was also true when the HVC lesioned and Control were grouped. The full statistical model that included Treatment (two lines NCM vs HVC + Control to fit; Fig. 6*B*) was not statistically distinguishable from the single line model that excluded Treatment as a predictor [$\chi^2(4) = 8.3925$, $p = 0.07{,}821$].

Similar results regarding the effect of NCM lesions on the previously learned stimuli but not on novel stimuli can be reached by examining the distribution across birds of the $\log_2$(OR) obtained for the first $k \leq 10$ uninterrupted trials (Fig. 5*C*,*F*). During this period and for the previously learned vocal signatures (Fig. 5*C*), each group scored significantly above chance level, but the mean $\log_2$(OR) was significantly smaller for the NCM lesioned birds [Wald test on the coefficient of mixed effect generalized statistical model testing for $\log_2$(OR) significantly different from zero; NCM, $\log_2$(OR) = 1.69 ± 0.342, $Z = 4.945$, $p < 0.001$; HVC, $\log_2$(OR) = 3.93 ±

0.559, $Z = 7.03$, $p < 0.001$; Control, $\log_2(OR) = 3.70$, $Z = 7.201$, $p < 0.001$, HVC + Control, $\log_2(OR) = 3.83$, $Z = 9.8$, $p < 0.001$]. Likelihood ratio test of mixed effect generalized models that included versus excluded treatment showed that the strength of the performance depended on lesion condition [$\chi^2(4)=15.936$, $p = 0.0031$], and a *post hoc* test showed that the NCM group did significantly worse than the combined HVC + Control group [difference in $\log_2(OR) = 2.12 + -0.58$, $Z = 4.092$ $p < 0.001$]. For the novel stimuli (Fig. 5*F*), each group also scored significantly above chance, but we failed to detect significant differences across Treatment [Wald test for $\log_2(OR) > 0$; NCM, $\log_2(OR) = 0.85$, $Z = 5.674$, $p < 0.001$; HVC, $\log_2(OR) = 1.24$, $Z = 6.536$, $p < 0.001$; Control, $\log_2(OR) = 1.30$, $Z = 9.658$, $p < 0.001$; HVC + Control, $\log_2(OR) = 1.26$, $Z = 9.776$, $p < 0.001$]. Likelihood ratio test of mixed-effect generalized models that included versus excluded treatment did not find statistical differences in the strength of the performance across lesion condition [$\chi^2(4) = 6.907$, $p = 0.1409$]. However, the effect of treatment became significant when HVC and Control data were combined; the NCM group performed slightly worse than the combined HVC + Control group [differences in $\log_2(OR) = 0.41 \pm 0.196$, $Z = 2.109$, $p = 0.035$].

Our data suggest that task performance of NCM lesioned birds in the early testing period was worse than in Control or HVC lesioned birds. However, performance on novel exposure appeared to be less affected than performance on previously experienced stimuli. The size of this effect can be quantified by a Cohen's d' measure that estimates the difference in the learning curves of NCM lesions and Control combined with the HVC lesioned birds in units of SD. These effect sizes are shown as a function of uninterrupted trials on Figure 6; the mean effect size for $k \leq 10$ is 2.74 ± 0.42 (2 SE) for previously learned stimuli and 0.86 ± 0.53 (2 SE) for novel stimuli. These results suggest that NCM lesioned birds lost the advantages birds acquired from previous exposure to the sounds but retained their ability to learn to classify new sounds (but see the section entitled "NCM lesions affect the maximum discrimination performance achieved after sustained learning" for caveats). Note that the postlesion learning curve of the NCM group for S1 vocal signatures (Fig. 5*A*,*B*) resembles the initial learning curve on naive stimuli before lesion (Fig. 4*A*) or the learning curve obtained for S2 vocal signatures after lesion (Fig. 5*D*,*E*), whereas the curves of the HVC and Control group resemble the curve measured late in learning (Fig. 4*B*). Examination of the intercepts and slopes of all linear fits obtained from mixed-effects models yielding the coefficients of log(OR) versus *k* learning curve validate these observations (Table 4).

Next, we investigated whether a more direct measure of an effect on the memory trace could be observed in NCM lesioned birds. Given that birds learn this task rapidly, we examined the performance before the very first uninterrupted trial experienced after lesions for S1. We determined the number of postlesion uninterrupted trials for which NoRe vocal signatures were interrupted from there on at significantly higher rates than Re vocal signatures. Because a measure based on statistical significance depends on sample size, and we had shown that HVC lesioned and Control subjects had indistinguishable postlesion learning curves, we compared NCM lesioned birds to HVC lesioned birds and Control birds grouped together. NCM lesioned subjects failed to distinguish among learned vocal signatures before the first postlesion uninterrupted trial, [Fisher's exact test on interruption counts across all vocalizers and subjects testing for $\log_2(OR) > 0$, $k = 0$, $\log_2(OR) = -0.1$, $p = 0.827$, Fisher's exact test per subject, 2/10 subjects with $p < 0.05$]. In contrast, the combined group of HVC and Control subjects performed the task significantly above chance before a single uninterrupted trial [Fisher's exact test on interruption counts across all vocalizers and subjects testing for log2(OR) > 0, $k = 0$, $\log_2(OR) = 2.01$, $p < 0.001$; Fisher's exact test per subject, 6/11 birds]. In the NCM lesioned birds, the $\log_2(OR)$ became significantly greater than zero from $k = 1$ uninterrupted trials onward. In the HVC lesioned and Control group, the $\log_2(OR)$ was significant from $k = 0$ onward.

NCM lesions affect the maximum discrimination performance achieved after sustained learning

We asked whether the steady-state performance reached prelesion levels after birds were given the opportunity to relearn. To answer this, we measured the average performance over the course of two entire sessions of 6v6-d2/8v8-d2 postlesion and compared it with the same period prelesion for the S1 stimulus set. In the postlesion period, all groups scored significantly above chance, demonstrating that they could do the task successfully after the operation [one-tailed one-sample *t* tests for $\log_2(OR) > 0$, where the OR is obtained for each bird from overall interruption counts [NCM, $\log_2(OR) = 2.11$, $t_{(9)} = 4.99$, $p < 0.001$; HVC, $\log_2(OR) = 4.03$, $t_{(6)} = 9.27$, $p < 0.001$; Controls, $\log_2(OR) = 4.12$, $t_{(3)} = 6.61$, $p = 0.004$]. On the one hand, and as we had also shown with the novel S2 stimuli, NCM lesioned birds retained the ability to learn or relearn this auditory discrimination task. On the other hand, their steady-state performance as measured on the second day [$\log_2(OR) = 2.11$] was worse than for the HVC [$\log_2(OR) = 4.03$] and Control birds [$\log_2(OR) = 4.12$]. To account for the potential random effect of bird, we compared the postlesion scores with prelesion scores of the difference in $\log_2(OR)$ between these sessions. NCM lesioned subjects performed more poorly after lesion [difference in $\log_2(OR) = -1.77$; $t_{(9)} = -3.60$, $p = 0.006$], HVC lesioned subjects did slightly worse [difference in $\log_2(OR) = -0.5$; $t_{(6)} = -1.94$, $p = 0.1$], and Controls improved slightly [difference in $\log_2(OR) = 0.83$, $t_{(3)} = 12.43$, $p = 0.001$]. Thus, NCM lesioned birds were not able to recover their prelesioned performance after having had a chance to relearn. One should note, however, that for S1 the prelesion training period is longer than the after lesion training as it includes an additional day of four versus four training. Thus, the poorer performance observed in NCM lesioned birds in contrast with the HVC and Control birds could be because of the loss of the memory acquired in the prelesion sessions. The poorer performance in NCM lesioned birds after lesion in contrast to their performance prelesion could be explained by a difference in learning regiment if one assumes they had indeed lost all auditory memories formed before the lesion. Alternatively, NCM lesioned birds might also have a more difficult time storing new memories. Examination of the performance for the S2 stimuli provides further insights into potential mechanisms.

For the novel stimulus set S2 used after lesion, we had failed to detect a significant difference in learning rates as a function of lesion treatment as described above. However, we noticed the NCM lesioned performance over the 10 first uninterrupted trials was slightly worse, reaching statistical significance only when we combined the HVC and Control groups (see analyses above; Figs. 5*B*, bottom row, 6*B*,*C*). Thus, we also examined the steady-state performance achieved by comparing the OR scores measured during the first 50 uninterrupted trials in the very last day of testing, the 6v6-d2/8v8-d2 sessions of S2. Fifty trials were chosen because the maximum number of uninterrupted trials experienced by a single bird on day 2 of testing was 58. As expected from the results already described for the first 10 uninterrupted

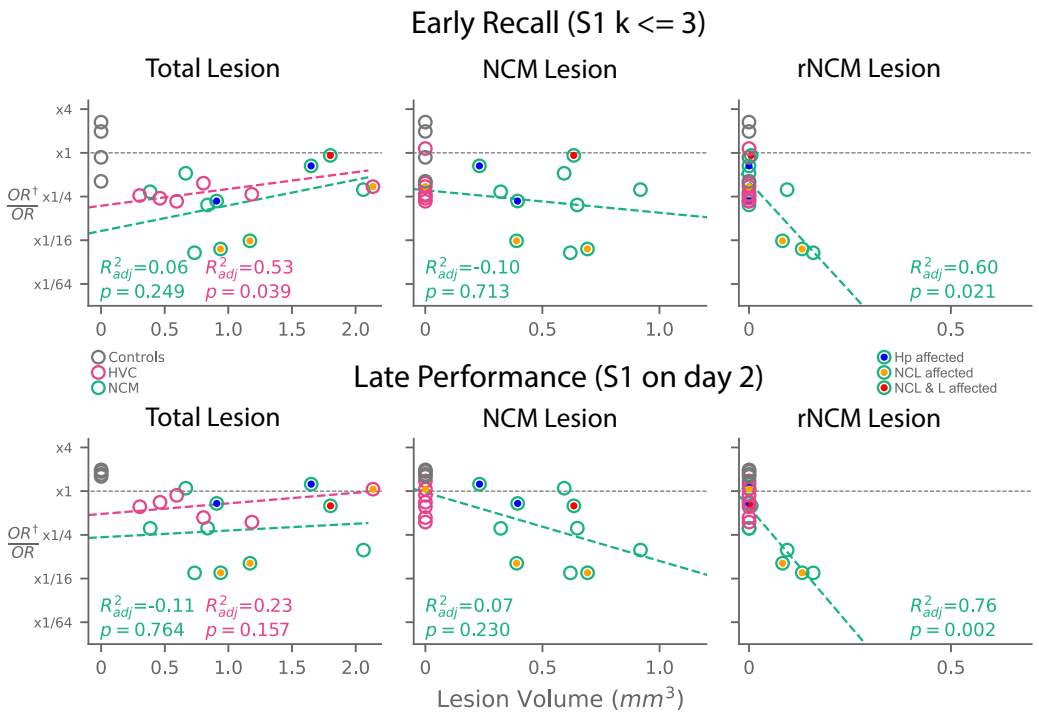

Figure 7. Task performance and lesion volume. The ratio of task performance after lesion ($OR^{\dagger}$) to before lesion (OR) is plotted as a function of the volume of the lesions. Left, The relationship for the total volume of the lesions both for the NCM- and HVC-targeted birds. Middle, The relationship for the lesion volume that was estimated to be in NCM proper for the NCM-targeted birds. Right, The relationship for the lesion volume found in the rostral quadrant of NCM (rNCM, see above, Materials and Methods; Fig. 2) for the NCM-targeted birds. Each point represents one subject. Values below one indicate worse performance after lesion. Top row, The ORs after lesion were calculated up to the first three uninterrupted trials to assess the effect of lesion size on the early recall. Bottom row, The ORs after lesion are estimated on d2. The ORs before lesion are always estimated on d2. Colors indicate lesion group (NCM, HVC, Controls) as used in Figures 3-6. Color matched dashed lines shows best fit line to data in the HVC and NCM group. Bottom left, Adjusted *R*-squared and *p* values for these fits are reported. The *p* values for the NCM quadrant analyses (right column) are Bonferroni corrected for multiple comparisons (i.e., multiplied by 4). Note the different scales used for the lesion volume axis (*x*-axis).

trials, the performance in all groups was significantly above chance [Fisher's exact test base on interruption counts for $\log_2(OR) > 0$; NCM, $\log_2(OR) = 2.26$, $p < 0.001$ and 9/10 birds with $p < 0.05$; HVC, $\log_2(OR) = 4.01$, $p < 0.001$ and 7/7 birds with $p < 0.05$; Controls, $\log_2(OR) = 3.54$ and 4/4 birds with $p < 0.05$]. This again shows that fully intact HVC and NCM are not required for zebra finches to remember novel sets of vocal signatures. However, we did find that overall performance on S2 after lesion for the NCM group was lower than that of HVC and Controls. This difference reached statistical significance when the NCM and HVC groups were combined [Wald test, difference in $\log_2(OR) = -1.74 \pm 0.68$, $Z = -3.68$, $p = 0.015$]. Although NCM lesioned birds were capable of relearning and with initial learning curves that were very similar to those of controls or HVC lesioned birds, their task performance measured after more extensive training was diminished; here, given equal opportunity to learn novel stimuli, NCM lesioned birds performed worse than HVC lesioned or Control birds. This decrease in performance could be from deficits in forming strong or stable auditory memories or, alternatively, other auditory perceptual deficits.

Relationship between NCM lesion volume and task performance

As described above, we observed an overall decrease in task performance in the NCM group relative to the HVC and Control groups when retested on S1 after lesion compared with before lesion. However, we also observed a large variability across subjects in each of these groups (Fig. 5*C*, top). Thus, we asked whether variability in the performance among the NCM or HVC subjects could be caused by variability in lesion size or exact location. NCM is a large brain region (Vates et al., 1996) that required several injections at multiple injection sites. HVC is a well-delimited nucleus, but its complete lesion also required multiple injections. Our lesion procedures thus resulted in variance in the total amount of brain damage both for the NCM- and HVC-targeted birds. To quantify the size of the lesions, we estimated the total volumetric extent of lesioned brain areas in Nissl-stained sagittal slices (Figs. 1*C*,*D*, 2). For the NCM and HVC groups, our lesions sometimes extended laterally to NCL or rostrally to field L (Extended Data Table 1-3). Finally, as NCM is a large region, we also divided it using a line parallel to L2 as a major axis and a second perpendicular line approximately at the dorsal/ventral midpoint (Fig. 2). In this manner, we obtained four quadrants that we labeled dorsal, rostral, ventral, and caudal. We estimated the volume of the lesions in each of these quadrants.

As shown in Figure 7, the overall lesion size (left columns) was not correlated with the performance in the behavioral task for either the NCM or HVC groups. The volume of the lesion restricted to NCM was not correlated with performance in early recall but a showed a trend for a correlation with a decrease in performance for the d2 sessions, which included a higher number of trials. This negative correlation between lesion size and performance became greater (and statistically significant) when the lesion volume was restricted to the rostral quadrant of NCM; both early recall and later performance were more affected in birds that had lesions within this rostral region. The number of birds with appreciable rostral NCM lesions is admittedly small. Note, however, that similar correlations and significances were obtained when the behavioral data were analyzed separately for song and distance calls (see below; Figs. 8-10) providing further

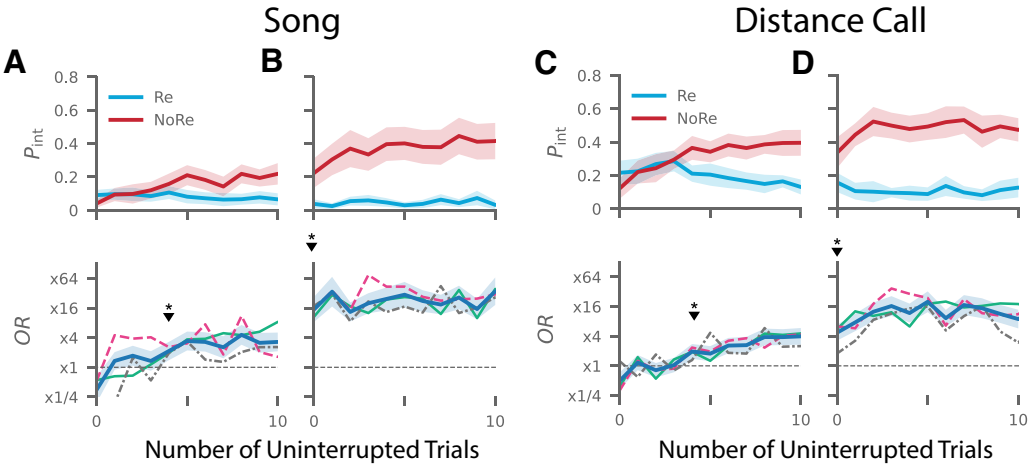

Figure 8. Prelesion S1 learning curves and steady-state performance split by Song/DC. ***A***, Probability of interruption (top) and odds ratio (bottom) between NoRe Song (red) and Re Song (blue) vocal signatures during initial exposure to S1, as in Figure 4. Asterisks show the first uninterrupted trial where OR was significantly higher than one from there on. ***B***. Late sessions of S1 8v8-d2 for Song, as in Figure 4*B*. Shaded regions in all subfigures indicate 2 SEM. ***C*** as in ***A***, but for DC. ***D*** as in ***B*** for late sessions of S1 6v6 d2 for DC.

evidence on the robustness of this effect. Also, as shown in Fig. 7, some of the lesions in HVC and NCM extended beyond their target to the hippocampus, NCL, and field L. Although sample sizes are small, none of these of target lesions resulted in additional decreases in performance. Two of the birds with significant lesions in rostral NCM had NCM lesions that extended beyond the NCM border (defined here at 1100 μm from the midline) into NCL but those two birds exhibited a similar task performance to the two birds that had similar NCM and rostral NCM lesion sizes without overflow into NCL.

Songs versus distance calls

In the preceding analyses, Song and DC stimuli were combined to compute $\log_2$(OR) scores. Songs and distance calls have different acoustical properties and serve different behavioral roles (Zann, 1996; Elie and Theunissen, 2016); memories used for individual recognition for each call type may use different neural circuits (Simpson and Vicario, 1990). Thus, we also measured task performance before and after lesion for Song and DCs separately. Although we observed some differences in the way birds treat the two vocalization types in our operant task, the effects of lesion and effect sizes were similar.

One clear difference between Song and DC is that healthy birds in the task are biased to interrupt DC at a higher probability than songs, even before any reward information has been gained (Yu et al., 2020). Over all the subjects ($n = 21$), birds interrupted songs with a probability of 0.07 before a single uninterrupted trial during initial learning of S1 vocal signatures (i.e., $k = 0$), which was significantly less than 0.17 for distance calls (two-tailed paired t test on p_{int} for DC vs Song; $t_{(20)} = 3.11$, $p = 0.006$). The apparent bias toward interrupting DCs persisted late in learning, as measured during the entirety of the 6v6-d2/8v8-d2 sessions (Song, $p_{int,Re} = 0.09$, $p_{int,NoRe} = 0.62$; DC, $p_{int,Re} = 0.18$, $p_{int,NoRe} = 0.74$).

The number of uninterrupted trials it takes prelesioned birds to distinguish Re from NoRe vocal signatures was similar for both call types [$k = 5$ for Song, $k = 4$ for DC for log(OR) significantly different from zero from there on; Fig. 8], and we verified that before lesion, memory for both Song and DC vocal signatures persisted over days/sessions, shown by higher than chance performance before the first uninterrupted trial on 6v6-d2/8v8-d2 sessions [one-tailed one-sample t test for $\log_2(OR) > 1$ with $k = 0$; Song, $\log_2(OR) = 3.61$, $t_{(20)} = 5.87$, $p < 0.001$; DC, $\log_2(OR) = 1.70$, $t_{(20)} = 2.12$, $p = 0.024$].

As we did before, the early responses to playbacks from distinct vocalizers after lesion were summarized by estimating $\log_2$(OR) for $k \leq 10$; we found that a mixed-effects model with additional intercepts per vocalization type (Song/DC) and interaction terms between vocalization type and lesion group did not explain the data better than a base model with just lesion group (likelihood ratio test;, $F_{(3,36)} = 2.017$, $p = 0.129$). Simply put, subjects were successful in this initial window after lesion for both call types, but NCM lesioned subjects did worse than the other groups for both (Fig. 9*A*,*B* for Song; Fig. 9*E*,*F* for DC; Table 5). The intercept obtained by fitting the learning curves with mixed-effect models for the NCM group was systematically lower than in the HVC group (Wald test, Song, $Z = -2.861$, $p = 0.0042$; DC, $Z = -3.952$, $p < 0.001$), Controls (Song, $Z = -1.75$, $p = 0.086$; DC, $Z = -1.72$, $p = 0.0848$), or combined HVC and Controls (Song, $Z = -2.93$, $p = 0.0028$; DC, $Z = -3.47$, $p = 0.005$), whereas there was no difference between HVC and Controls (Song, $Z = 0.976$, $p = 0.329$; DC, $Z = 1.24$, $p = 0.2161$). Similarly, the performance measured by the log(OR) for $k \leq 10$ for the S1 postlesion was lower in NCM group than in the combined HVC + Control group both for Song (unpaired two-sided t test, $t_{(30)} = 2.35$, $p = 0.029$) and for DC ($t_{(30)} = 3.47$, $p = 0.003$). Finally, we found that modeling the change in steady-state (6v6-d2/8v8-d2) performance for S1 after lesion compared with before lesion was not improved by including terms for vocalization type and the interaction between vocalization type and lesion group ($F_{(3,36)} = 0.589$, $p = 0.626$).

We repeated our analysis of S2 performance for Song and DC separately (Fig. 9*C*,*D* for Song, *G*,*H* for DC). In the learning phase, for uninterrupted trials up to $k = 10$, we verified that just as for the combined data, a mixed-effects linear model of $\log_2$(OR) with three lines (one for each lesion group) was not a statistically better fit than a model with a single line. This was true for both vocalization types [likelihood ratio test, Song, $\chi^2(8) = 11.354$, $p = 0.1824$; DC, $\chi^2(8) = 12.752$, $p = 0.1207$]. At the end of learning, overall scores for S2 on sessions 6v6-d2/8v8-d2 were above chance level for Song and DC. We tested the overall $\log_2$(OR) score during these sessions for S2 using a linear mixed-effects model of $\log_2$(OR), with lesion group as a fixed effect and subject identity as a random effect. We compared this

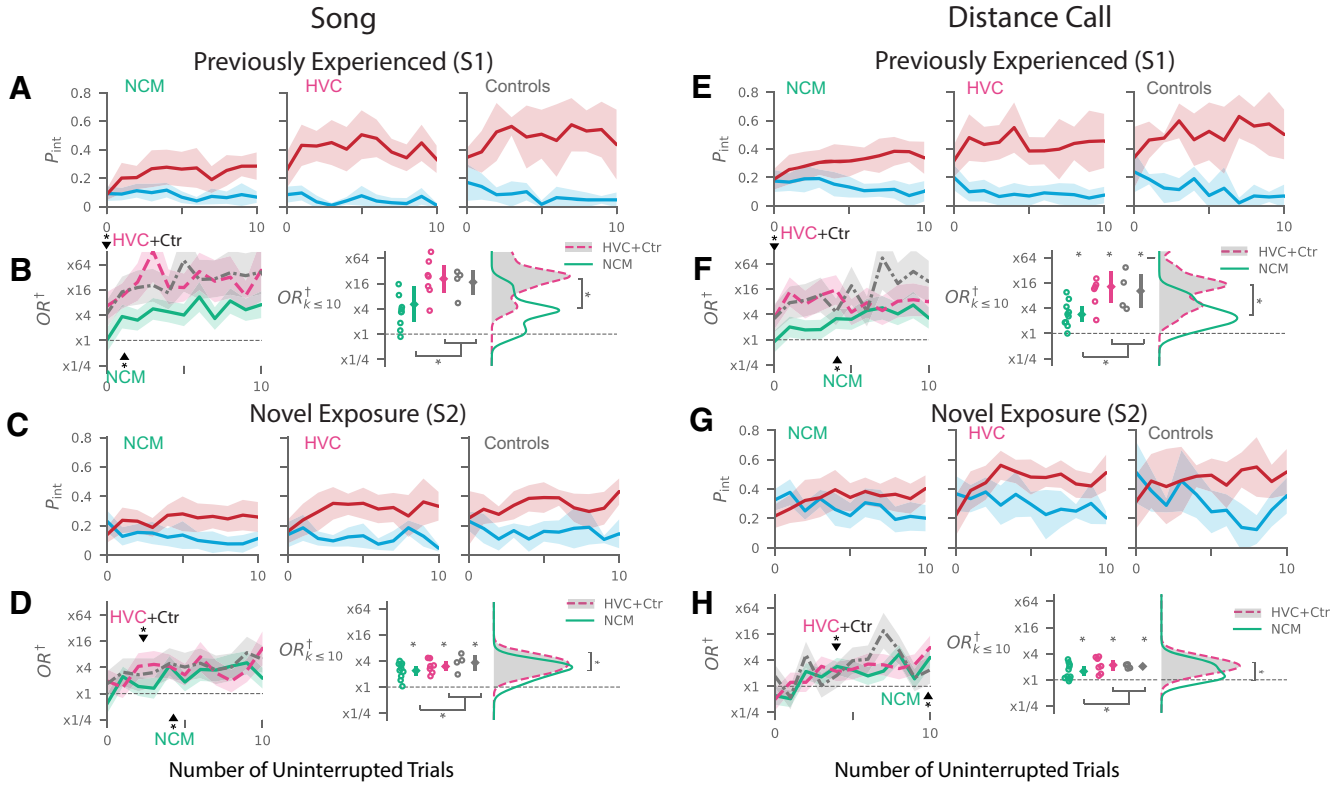

Figure 9. Postlesion learning curves for previously experienced and novel stimuli split by Song/DC. ***A***, Probability of interruption for NoRe and Re vocal signatures in the immediate session after lesion for S1, as in Figure 4*A* but for Song stimuli only. ***B***, Odds ratios for three lesion groups, NCM, HVC, and Controls, in the immediate sessions after lesion for S1, as in Figure 5*B*, but for Song stimuli only. Right, scatter plot shows the distribution of OR obtained in this early exposure period estimated for $k \leq 10$. ***C***, ***D***, Like ***A*** and ***B*** but for the initial experience with the novel stimulus set after lesion, S2. Significance stars above the scatter plot indicate one-sided one-sample *t* test results for $\log_2 OR^{\dagger} > 0$ with significance threshold 0.05. For the previously experienced S1 for Song (***B***) NCM, $\log_2 OR = 2.21$, $Z = 3.737$, $p < 0.0002$; HVC, $\log_2 OR = 4.31$, $Z = 8.604$, $p < 0.0001$; Controls, $\log_2 OR = 4.11$, $Z = 7.894$ $p < 0.001$. Shaded regions show 2 SEM. Below bar and stars show significance of difference in $\log_2 OR$ between NCM and HVC + Controls combined, $\delta \log 2OR = -2.16$, $Z = -3.351$, $p = 0.008$. For the novel exposure of S2 vocal signatures based on song (***D***), NCM, $\log_2 OR = 1.24$, $Z = 6.409$, $p < 0.0001$; HVC, $\log_2 OR = 1.56$, $Z = 7.804$, $p < 0.0001$; Controls, $\log_2 OR = 1.86$, $Z = 6.508$, $p < 0.0001$. Star below the $\times 1$ dashed line in the scatter plot in ***C*** shows in $\log_2 OR$ between NCM and HVC + Controls combined, $\delta \log 2OR = -0.52$, $Z = -2.091$, $p = 0.0365$. ***E–H***, Like ***A–D*** but for DC. For the previously experienced S1 for DC (***F***), NCM, $\log_2 OR = 1.47$, $Z = 4.809$, $p < 0.0001$; HVC, $\log_2 OR = 3.55$, $Z = 6.319$, $p < 0.0001$; Controls, $\log_2 OR = 3.34$, $Z = 4.841$, $p < 0.001$. Star below the $\times 1$ dashed line in the scatter plot in ***F*** shows significance of difference in $\log_2 OR$ between NCM and HVC + Controls combined, $\delta \log_2 OR = -1.93$, $Z = 3.714$, $p = 0.0002$. For the novel exposure of S2 vocal signatures based on DC (***H***), NCM, $\log_2 OR = 0.62$, $Z = 3.436$, $p = 0.006$; HVC, $\log_2 OR = 1.12$, $Z = 5.175$, $p < 0.0001$; Controls, $\log_2 OR = 1.03$, $Z = 6.557$, $p < 0.0001$. Star below the $\times 1$ dashed line in the scatter plot in ***H*** shows significance of difference in $\log_2 OR$ between NCM and HVC + Controls combined, $\delta \log 2OR = -0.45$, $Z = 1.993$, $p = 0.0462$.

Table 5. Task performance split by songs and DCs during initial re-exposure to S1 after lesion

Call type	Group	df	Effect size ($\log_2 OR$)	*t* Statistic	*p* value
Song	NCM	9	2.36	3.21	0.011
Song	HVC	6	4.38	7.58	<0.001
Song	Controls	3	4.11	7.32	0.005
DC	NCM	9	1.50	4.78	0.001
DC	HVC	6	3.70	5.48	0.002
DC	Controls	3	3.36	4.33	0.023

Statistics on $\log_2(OR)$ for $k \leq 10$ during the initial re-exposure to S1 after lesion. Tests were one-tailed one-sample *t* tests to detect $\log_2(OR) > 0$, applied to Song and DC trials separately. OR ratios were based on total number of interrupted trials for each bird, call type, and condition.

to a model that additionally included an intercept for vocalization type (Song/DC) and interaction terms between vocalization type and lesion group. We found that including call type parameters better fit the data (likelihood ratio test, $F_{(3,36)} = 19.06$, $p < 0.001$) and that this was driven by higher performance on songs than DCs (Wald test, $Z = 2.27$, $p = 0.023$), whereas the interaction between vocalization type and lesion group was not significant (Wald test, HVC × Song, $Z = 1.44$, $p = 0.150$; NCM × Song, $Z = 0.642$, $p = 0.521$).

The effect of the size of the lesion and its location in the NCM group were also very similar for Song and DC. When comparing the reduction in performance (i.e., the ratio of OR after lesion to before lesion), the discrimination of vocal signatures based on Song and DC is equally affected and this reduction in performance is equally correlated with the size of the lesion in the rostral quadrant of NCM (Fig. 10).

In summary, lesions to NCM affect in a very similar manner the discrimination of vocalizer ID based on song versus DC even if we note some behavioral differences with these two call types; although Song triggers fewer interruptions, potentially because birds tend to want to listen to its entirety, it is also ultimately discriminated with slightly higher performance.

Task performance before and after lesion

To verify that lesioned birds were able to perform the discrimination task and that deficits in hearing or other cognitive functions were not also impaired (e.g., ability to interrupt, and task understanding), we assessed their performance on the easiest version of the task, where there was only one Re vocalizer and one NoRe vocalizer (1v1). This task has a low memory load as it can be performed by detecting low-probability stimuli as the rewarded ones. The average performance measured over the entire day for S1 1v1 before lesion (S1), S1 1v1 after lesion ($S1^{\dagger}$), and S2 1v1 after lesion ($S2^{\dagger}$) are shown on the top row of Fig. 11. Note that

although they all describe the easiest version of the task, S1 and S2† are sessions during which the stimuli were completely novel to the subjects, whereas in S1† the stimuli had been heard previously for several days before lesion. One might therefore expect the performance on S1† to be worse for NCM lesioned birds and better for controls and HVC lesioned birds. As shown in Fig. 11 (top row), on average, the birds performed very well on the 1v1 task in every condition with average OR between 4 and 64. In particular, the average performance of NCM lesioned birds for S1† postlesion was not statistically different from that observed for S1 prelesion, even if one (for Song) and two (for DC) birds performed worse than before the lesion. The performances on this low memory-load task should be compared with the performance on the high memory-load task achieved on day 2 for the 6v6 and 8v8 discriminations (Fig. 11, bottom row). Here, the average decrease in performance for the NCM lesioned birds is clear. Thus, NCM lesioned birds that could clearly perform well in the 1v1 task (i.e., those beyond the two exceptions) showed significant decreases in performance in the 6v6 and 8v8 tasks, the tasks that required a higher memory load.

HVC lesion effect on song production

HVC lesions are known to have acute effects on song production in songbirds (Nottebohm et al., 1976; Aronov et al., 2008). In addition to histologic verification of HVC lesions (Fig. 1*D*), we compared songs produced before and after the experiment as a secondary way to validate HVC lesion completeness. Of the five male HVC group subjects, all five produced normal songs before the experiment, but only two birds produced songs after the experiment. Both subjects produced songs with substantial spectral and temporal degradations (Fig. 12), consistent with HVC ablation.

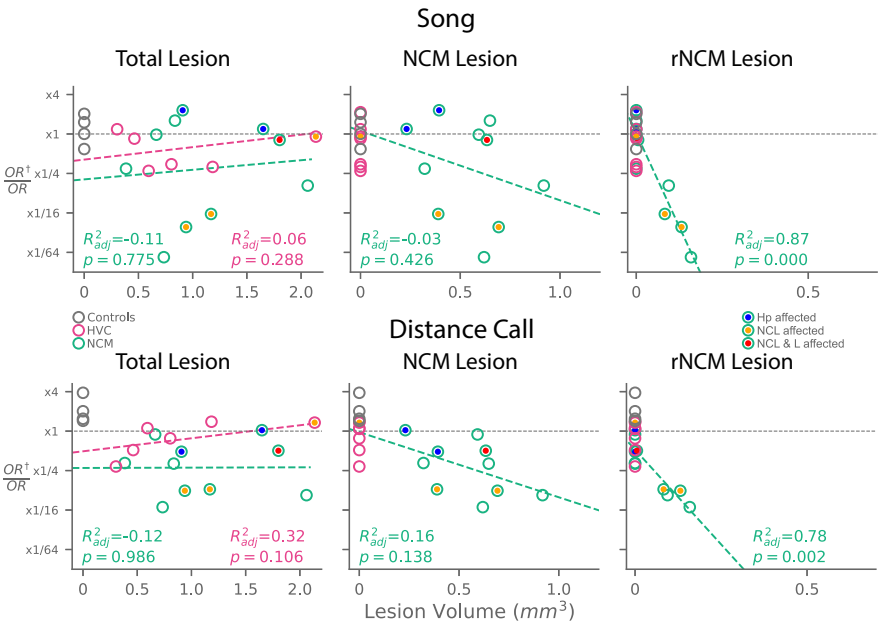

Figure 10. Task performance and lesion volume split by Song/DC. As in Figure 7, the ratio of task performance after lesion (OR†) to before lesion (OR) for S1 on d2 (Late Performance) is plotted as a function of estimated total lesion volume, NCM lesion volume and rostral NCM (rNCM) lesion volume for Song (top row) and Distance Calls (bottom row). The *p* values for the rNCM are Bonferroni corrected.

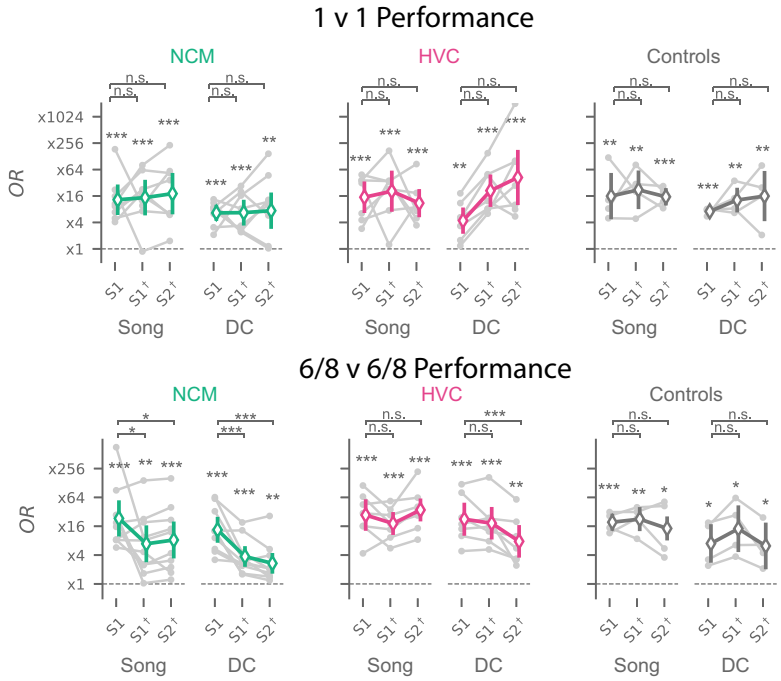

Figure 11. Task performance on 1v1 versus 6v6(DC) or 8v8 (Song) before and after lesion. The figure shows the OR estimated over the entire 1v1 day (top row) or the entire day 2 for 6v6 (DC) or 8v8 (Song) before and after lesions and for S1 and S2. Significance of one-tailed one-sample *t* tests for $\log_2(OR) > 0$ shown above plot with isolated stars; significance one-tailed paired *t* test for difference in log(OR) between postlesion and prelesion < 0 (decrease in performance) shown with brackets; ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, n.s. Note that the paired comparison between S2† and S1 was performed with different stimuli. Differences in performance in the 6v6, 8v8 for S2† versus S1 might also be because of differences in the difficulty of discriminating stimuli within each of these ensembles.

Discussion

We established a role of the secondary auditory pallial area, NCM, in the ability of the zebra finch to store and recall auditory memories for the identifying features in the song and distance calls of distinct vocalizers. We found that performance in a memory task was impaired immediately after NCM lesion, as measured by the inability of the bird to distinguish previously memorized rewarded and nonrewarded vocal signatures. NCM lesioned birds retained the ability to form novel memories, evidenced by early learning rates that were similar to prelesion rates and steady-state performance above chance. The steady-state performance after lesion was, however, also below the prelesion levels. Both the reduction in the immediate recall performance and the reduction in the ability to form multiple new auditory memories were correlated with the size of the lesion in

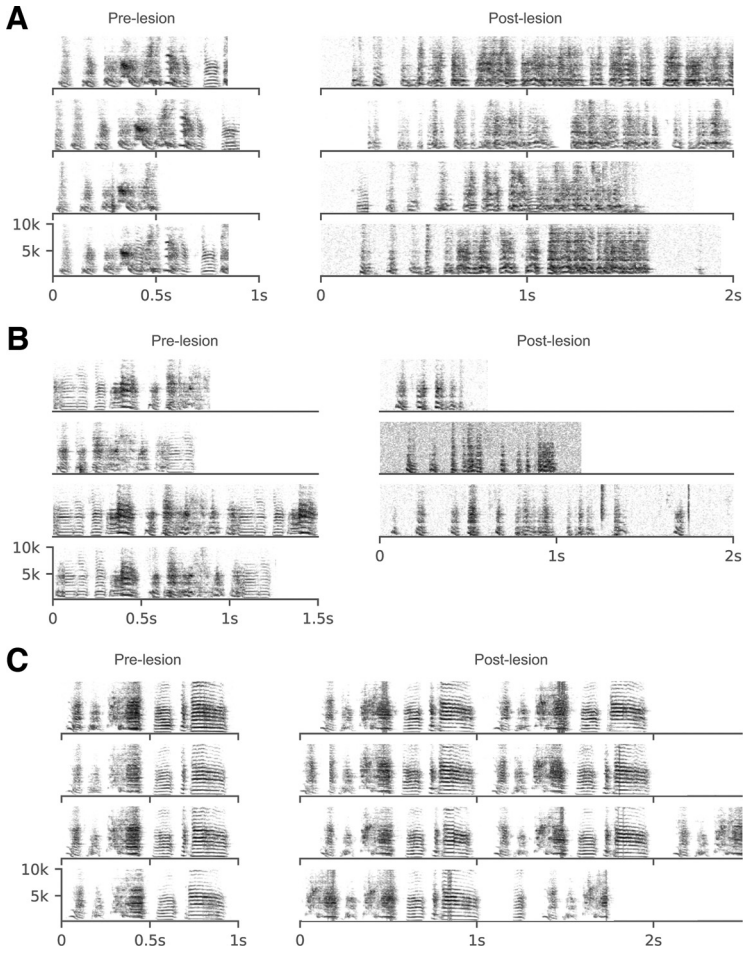

Figure 12. Validation of song degradation following HVC lesion. We validated HVC lesions by comparing songs performed before and after the experiment. ***A***, ***B***, Of the five males in the HVC lesioned group, three males did not sing postlesion. These figures show spectrograms of song motifs produced by the two males that sang postlesion, both prelesion (left) and postlesion (right). Nissl-stained histology of HVC lesion for ***A*** shown in Figure 1*D*. ***C***, Spectrograms of songs produced before (left) and after (right) lesions in a male assigned to the Control group. Bilateral injections of ibotenic acid were made to HVC coordinates (Extended Data Table 1-1), but histology showed that the lesion was ultimately off target, and song production postlesion was not noticeably affected.

rostral NCM. These effects were observed both for the recognition of vocalizers based on song and based on distance calls. In contrast, birds with lesions to the premotor nucleus HVC showed deficits in vocal production but performed as well as controls in the auditory memory recall task during the initial postlesion trials as well as in their steady-state performance for discrimination of novel vocal signatures after lesion.

Zebra finches only required a few trials before performing significantly above chance level in this auditory memory task. This rapid learning made it challenging to detect any lesion effect on previously learned memories, as task performance can quickly reflect not only recall of previously learned stimuli but also further reinforcement of these memories or the learning of new ones. To address this strong testing effect (Roediger and Butler, 2011), we rigorously quantified learning curves as a function of uninterrupted trials, the only trials that provided feedback to the birds regarding reward contingency. This approach allowed us to separate potential effects of our lesions on relearning from the effects on recall, which could only be detected at the very beginning of the testing sessions. After bilateral NCM lesions, there is a remarkable behavioral contrast between the apparent degradation of previously learned discriminations and preservation of learning and relearning abilities (Figs. 3, 4, 5).

NCM is a large brain region without well-defined boundaries, and it is therefore very difficult to lesion the entire nucleus. Our injections, although large and bilateral, resulted in lesions of variable sizes (Fig. 6). It is therefore possible that areas in NCM that remained intact provide the neural circuits needed for the formation of novel memories. We also found that although the learning curves for novel vocal signatures following lesion are almost identical to those found before the lesion, the ultimate steady-state performance for the tasks with high memory load (8v8 songs and 6v6 DC) was affected. We also found that both the recall of prior memories and the ability for relearning were correlated with the size and location of the lesions; the birds with large lesions in the rostral region of NCM had the poorest performance. These results suggest that NCM (particularly rostral NCM) plays an important role for storing memories; its destruction prevents recall of previously learned memories and affects the ultimate performance reached for novel associations in this complex auditory memory task. Alternatively, lesions to NCM could have affected auditory perception making the discrimination task more difficult. Although we can't rule out this possibility, the fact that early recall can be abolished, whereas relearning and novel learning can still occur, suggests than NCM lesions have a stronger effect on the memory trace than on auditory discrimination. We also found that the performance for tasks with smaller memory load were less affected (Fig. 11), further suggesting that putative auditory discrimination effects alone could not fully explain these results.

The preserved ability for relearning and novel learning in NCM lesioned birds could also be mediated by other auditory pallial areas. The two major avian secondary auditory areas NCM and CM are interconnected (Vates et al., 1996), and neurophysiological recordings performed after auditory discrimination tasks have shown neural correlates of auditory memories in CM (Gentner and Margoliash, 2003; Jeanne et al., 2011). It is also known that the primary auditory areas that are presynaptic to CM and NCM can generate invariant responses to different renditions of conspecific vocalizations (Meliza and Margoliash, 2012). Finally, one could postulate that yet higher-level pallial areas, such as NCL, which is thought to contain more abstract representations for goal-directed action (Rinnert and Nieder, 2021) and neural correlates of working memory (Rinnert et al., 2019), could also be involved in the memory functions required for this go/no-go task. The formation and storage of auditory memories and their recall almost certainly involves a distributed network in the avian pallium.

Within this distributed network, NCM appears to play a central role for storing auditory memories. We have provided direct evidence for this role, complementing multiple prior studies that described various neural correlates of plasticity and memory for song in NCM. It has been well described that NCM is a site of strong expression of immediate early genes implicated in

plasticity and that this expression can be simply triggered by song exposure (Mello et al., 1992; Mello and Clayton, 1994). Repeated presentations of the same song result in stimulus-specific adaptation in the neurophysiological recordings, and this adaptation requires protein synthesis (Chew et al., 1995; Mello et al., 1995). NCM has also been implicated in multiple studies in the storage of the tutor song memory that young male birds must form in the context of song learning (London and Clayton, 2008; Gobes et al., 2010; Yanagihara and Yazaki-Sugiyama, 2016; Katic et al., 2022). Our study shows that NCM is involved in storing auditory memories of songs and calls of other conspecifics not for the purpose of imitating these sounds but instead for recognizing the vocalizer. Thus, we suggest that NCM acts as a general purpose high-level auditory region specialized for the recognition and memorization of complex sounds including the vocalizations of conspecifics that are behaviorally relevant.

However, the exact role of NCM in the formation and recall of auditory memories is yet unclear, and many questions remain. For example, one set of loss-of-function studies in NCM targeting the song imitation behavior has produced mixed results; during song learning, lesions to NCM in juvenile birds do not appear to inhibit song learning, although those lesions might not have been complete (Canopoli et al., 2016, 2017). It should also be noted that strong causal evidence for putative memory traces of the tutor song have been described in many other brain regions including the Interfacial (Nif), Avalanche (Av) and HVC song nuclei in the pallium (Roberts et al., 2017; Zhao et al., 2019) and midbrain dopaminergic centers (Tanaka et al., 2018). Thus, although NCM might play a role in storing the template for the song imitation behavior, just as we postulate a distributed network for storing and recalling the auditory memories required for vocalizer recognition, the storing of the tutor song template has been shown previously to be dependent on multiple brain regions including areas in the avian song system, basal ganglia, and ventral tegmental area (for review, see Sakata and Yazaki-Sugiyama, 2020).

Multiple studies have also begun to examine the neuromodulatory mechanisms, and in particular the role of estrogen, for facilitating neural plasticity in NCM and consequently song learning. For example, suppression of aromatase activity in NCM has been shown to inhibit auditory association learning but not to affect recall of previously learned stimuli (Macedo-Lima and Remage-Healey, 2020). Surprisingly, unilateral estrogen synthesis blockade did not affect the quality of learned songs but instead led to enhanced neural responses to tutor song in NCM and HVC (Vahaba et al., 2020). A particularly interesting target for future manipulations is the dopaminergic innervations of NCM, which have been shown to have effects on learning of auditory stimuli and preference to songs (Chen et al., 2016; Barr et al., 2021; Macedo-Lima et al., 2021). Identifying and manipulating a reward signal projecting to NCM, if it exists, could help to disambiguate between memory for individual recognition and memory for reward associations. Such a study would benefit from acute, reversible manipulations as opposed to the chronic, irreversible lesions performed in the current study.

We found that lesions in rostral NCM were particularly deleterious to auditory memory performance or auditory discrimination. Rostral NCM is the principal recipient of the output projections originating from the primary auditory pallial region, L2a, to NCM (Vates et al., 1996). In seasonal birds, the conspecific song selectivity, which is estrogen dependent, has also been shown to be restricted to rostral NCM (Sanford et al., 2010). In our prior neurophysiological experiments, we had also shown that the more ventral regions of NCM have selective and invariant responses to song in the sense that they are robust in the presence of noise (Moore et al., 2013). The same region contains auditory neurons that have high information about the individual signature found in distance calls even when these are degraded as a result of natural propagation (Mouterde et al., 2017). A reexamination of those data show a fair amount of anatomic overlap with the rostral quadrant shown here to be so critical for the learned discrimination of vocalizer identification. Additional neurophysiological experiments targeted to this region are needed to further understand the nature of the auditory information represented in the neural responses and how it is affected by learning.

The memory task used in this experiment was designed to be naturalistic, based on the natural vocal behavior of the zebra finch as a gregarious species living in relatively large groups (Zann, 1996) and with multiple social behaviors requiring individual recognition (for review, see Elie and Theunissen, 2020). Our aim was to interrogate the role of NCM in auditory memory while this natural skill is being used by using a large number of renditions of songs and calls for each vocalizer. Despite these efforts to maximize the behavioral relevance of the categories of sounds being discriminated, the operant conditioning testing remains an artificial setting, and it is known that intensive training on sound stimuli using playbacks may influence perception, the underlying neural code, and even use different neural pathways (Bennur et al., 2013). Birds in our task were isolated during the sessions, whereas live social interaction has been shown to influence the quality of song learning in birds (Eales, 1989; Chen et al., 2016; Yanagihara and Yazaki-Sugiyama, 2019), and social reinforcement in the form of video playback of other birds can be sufficient motivation for learning in an operant task (Macedo-Lima and Remage-Healey, 2020). Thus, we must consider the possibility that individual vocal recognition in natural settings with social consequences may engage different pathways and areas of the brain than those tested in our experiment. Additional experiments that test the natural responses of these birds to known individuals may further illuminate the role of NCM and other auditory regions for auditory memory.

Unlike with NCM, we found little to no effect of HVC lesions on auditory memory in this recognition task. HVC is the critical interface between the auditory and song systems (Margoliash, 1997; Roberts and Mooney, 2013) and gates the auditory information needed to give rise to song-selective neurons found throughout the vocal production pathway and the anterior forebrain pathway for song learning (Vicario and Yohay, 1993; Roberts and Mooney, 2013). HVC lesions are used in the current experiment serve as a proxy to test the necessity of the vocal motor pathways on auditory memory for individual vocal recognition; we found no effect of HVC lesion during initial trials after lesion—which was indistinguishable from the normal behavioral response to learned vocal signatures in healthy subjects—nor in the overall task performance over entire sessions. Our HVC lesions did, however, result in serious deficits in song production as expected. Thus, we believe that a motor representation of conspecific communication sounds is not necessary for the basic perceptual, memory, and recognition functions required to succeed in this vocalizer recognition task. Previous studies had shown that HVC lesions can alter courtship behavior and mate preference in females (Brenowitz, 1991; Del Negro et al., 1998; Perkes et al., 2019) and the association between a song and a referent in males

(Gentner et al., 2000) but not the discrimination of song; similarly, our results suggest that HVC is not needed for the perceptual identification of the song of conspecifics. However, the song of the mate was not tested directly in our study. Thus, it is possible that HVC (and the song system) plays a role in modulating the specific mate-related behavioral responses once the perceptual recognition is achieved.

The auditory memory function in zebra finches in terms of its capacity and speed is remarkable and reminiscent of the fast learning ability shown in humans for auditory-based word learning (Markson and Bloom, 1997). Social birds that rely on vocal communication for creating and preserving social bonds might therefore share similar auditory capacities to those of vocal social primates, including humans. In both birds and primates, it is interesting to note that memory for behaviorally relevant sounds appears to depend on sensory cortical areas (Weinberger, 2004) and not necessarily traditional memory regions such as the hippocampus (Vargha-Khadem et al., 1997; Fritz et al., 2005). More specifically, the neural regions in primates that have been implicated in the recognition of individual vocal signatures, the voice regions, might be in a secondary auditory region found in the temporal lobe (Perrodin et al., 2011). This secondary auditory cortical region is analogous both anatomically and genomically to the avian secondary pallial regions NCM or CM (Jarvis et al., 2013). Thus, unraveling the neural circuits and neural mechanisms that underlie the formation and recall of these fast memories in social songbirds will provide unique insights into the function of high-level auditory areas for vocal communication requiring the recognition of learned auditory categories.

References

Aronov D, Andalman AS, Fee MS (2008) A specialized forebrain circuit for vocal babbling in the juvenile songbird. Science 320:630–634.

Barr HJ, Wall EM, Woolley SC (2021) Dopamine in the songbird auditory cortex shapes auditory preference. Curr Biol 31:4547–4559.e5.

Bennur S, Tsunada J, Cohen YE, Liu RC (2013) Understanding the neurophysiological basis of auditory abilities for social communication: a perspective on the value of ethological paradigms. Hear Res 305:3–9.

Bolhuis JJ, Gahr M (2006) Neural mechanisms of birdsong memory. Nat Rev Neurosci 7:347–357.

Bolhuis JJ, Hetebrij E, Den Boer-Visser AM, De Groot JH, Zijlstra GG (2001) Localized immediate early gene expression related to the strength of song learning in socially reared zebra finches. Eur J Neurosci 13:2165–2170.

Bolhuis JJ, Okanoya K, Scharff C (2010) Twitter evolution: converging mechanisms in birdsong and human speech. Nat Rev Neurosci 11:747–759.

Brenowitz EA (1991) Altered perception of species-specific song by female birds after lesions of a forebrain nucleus. Science 251:303–305.

Canopoli A, Herbst JA, Hahnloser RHR (2014) A higher sensory brain region is involved in reversing reinforcement-induced vocal changes in a songbird. J Neurosci 34:7018–7026.

Canopoli A, Zai A, Hahnloser R (2016) Lesions of a higher auditory brain area during a sensorimotor period do not impair birdsong learning. Matters 1–5.

Canopoli A, Zai A, Hahnloser RHR (2017) Bilateral neurotoxic lesions in NCM before tutoring onset do not prevent successful tutor song learning. Matters 12:20.

Carlson NV, Kelly EM, Couzin I (2020) Individual vocal recognition across taxa: a review of the literature and a look into the future. Philos Trans R Soc Lond B Biol Sci 375:20190479.

Chen Y, Matheson LE, Sakata JT (2016) Mechanisms underlying the social enhancement of vocal learning in songbirds. Proc Natl Acad Sci U S A 113:6641–6646.

Chew SJ, Mello C, Nottebohm F, Jarvis E, Vicario DS (1995) Decrements in auditory responses to a repeated conspecific song are long-lasting and require two periods of protein synthesis in the songbird forebrain. Proc Natl Acad Sci U S A 92:3406–3410.

D'Amelio PB, Klumb M, Adreani MN, Gahr ML, Ter Maat A (2017) Individual recognition of opposite sex vocalizations in the zebra finch. Sci Rep 7:5579.

Del Negro C, Gahr M, Leboucher G, Kreutzer M (1998) The selectivity of sexual responses to song displays: effects of partial chemical lesion of the HVC in female canaries. Behav Brain Res 96:151–159.

Eales LA (1989) The influences of visual and vocal interaction on song learning in zebra finches. Animal Behaviour 37:507–508.

Elie JE, Theunissen FE (2016) The vocal repertoire of the domesticated zebra finch: a data-driven approach to decipher the information-bearing acoustic features of communication signals. Anim Cogn 19:285–315.

Elie JE, Theunissen FE (2018) Zebra finches identify individuals using vocal signatures unique to each call type. Nat Commun 9:4026.

Elie JE, Theunissen FE (2020) The neuroethology of vocal communication in songbirds: production and perception of a call repertoire. In: The neuroethology of birdsong (Sakata JT, Woolley SC, Fay RR, Popper AN, eds), pp 175–209. Cham, Switzerland: Springer.

Eugen KV, Tabrik S, Güntürkün O, Ströckens F (2020) A comparative analysis of the dopaminergic innervation of the executive caudal nidopallium in pigeon, chicken, zebra finch, and carrion crow. J Comp Neurol 528:2929–2955.

Formisano E, De Martino F, Bonte M, Goebel R (2008) "Who" is saying "what"? Brain-based decoding of human voice and speech. Science 322:970–973.

Fritz J, Mishkin M, Saunders RC (2005) In search of an auditory engram. Proc Natl Acad Sci U S A 102:9359–9364.

Gentner TQ, Margoliash D (2003) Neuronal populations and single cells representing learned auditory objects. Nature 424:669–674.

Gentner TQ, Hulse SH, Bentley GE, Ball GF (2000) Individual vocal recognition and the effect of partial lesions to HVc on discrimination, learning, and categorization of conspecific song in adult songbirds. J Neurobiol 42:117–133.

Gobes SMH, Bolhuis JJ (2007) Birdsong memory: a neural dissociation between song recognition and production. Curr Biol 17:789–793.

Gobes SMH, Zandbergen MA, Bolhuis JJ (2010) Memory in the making: localized brain activation related to song learning in young songbirds. Proc R Soc B 277:3343–3351.

Hahnloser RHR, Kozhevnikov AA, Fee MS (2002) An ultra-sparse code underlies the generation of neural sequences in a songbird. Nature 419:65–70.

Jarvis ED, Yu J, Rivas MV, Horita H, Feenders G, Whitney O, Jarvis SC, Jarvis ER, Kubikova L, Puck AEP, Siang-Bakshi C, Martin S, McElroy M, Hara E, Howard J, Pfenning A, Mouritsen H, Chen C-C, Wada K (2013) Global view of the functional molecular organization of the avian cerebrum: mirror images and functional columns. J Comp Neurol 521:3614–3665.

Jeanne JM, Thompson JV, Sharpee TO, Gentner TQ (2011) Emergence of learned categorical representations within an auditory forebrain circuit. J Neurosci 31:2595–2606.

Katic J, Morohashi Y, Yazaki-Sugiyama Y (2022) Neural circuit for social authentication in song learning. Nature Communication 13:4442.

London SE, Clayton DF (2008) Functional identification of sensory mechanisms required for developmental song learning. Nat Neurosci 11:579–586.

Long MA, Fee MS (2008) Using temperature to analyse temporal dynamics in the songbird motor pathway. Nature 456:189–194.

Macedo-Lima M, Remage-Healey L (2020) Auditory learning in an operant task with social reinforcement is dependent on neuroestrogen synthesis in the male songbird auditory cortex. Horm Behav 121:104713.

Macedo-Lima M, Boyd HM, Remage-Healey L (2021) Dopamine D1 receptor activation drives plasticity in the songbird auditory pallium. J Neurosci 41:6050–6069.

Margoliash D (1997) Functional organization of forebrain pathways for song production and perception. J Neurobiol 33:671–693.

Markson L, Bloom P (1997) Evidence against a dedicated system for word learning in children. Nature 385:813–815.

Meliza CD, Margoliash D (2012) Emergence of selectivity and tolerance in the avian auditory cortex. J Neurosci 32:15158–15168.

Mello C, Nottebohm F, Clayton D (1995) Repeated exposure to one song leads to a rapid and persistent decline in an immediate early gene's response to that song in zebra finch telencephalon. J Neurosci 15:6919–6925.

Mello CV, Clayton DF (1994) Song-induced ZENK gene expression in auditory pathways of songbird brain and its relation to the song control system. J Neurosci 14:6652–6666.

Mello CV, Vicario DS, Clayton DF (1992) Song presentation induces gene expression in the songbird forebrain. Proc Natl Acad Sci U S A 89:6818–6822.

Moore RC, Lee T, Theunissen FE (2013) Noise-invariant neurons in the avian auditory cortex: hearing the song in noise. PLoS Comput Biol 9: e1002942.

Mouterde SC, Elie JE, Mathevon N, Theunissen FE (2017) Single neurons in the avian auditory cortex encode individual identity and propagation distance in naturally degraded communication calls. J Neurosci 37:3491–3510.

Nottebohm F, Stokes TM, Leonard CM (1976) Central control of song in the canary, *Serinus canarius*. J Comp Neurol 165:457–486.

Pagliaro AH, Arya P, Piristine HC, Lord JS, Gobes SMH (2020) Bilateral brain activity in auditory regions is necessary for successful vocal learning in songbirds. Neurosci Lett 718:134730.

Perkes A, White D, Wild JM, Schmidt M (2019) Female songbirds: the unsung drivers of courtship behavior and its neural substrates. Behav Processes 163:60–70.

Perrodin C, Kayser C, Logothetis NK, Petkov CI (2011) Voice cells in the primate temporal lobe. Curr Biol 21:1408–1415.

Phan ML, Pytte CL, Vicario DS (2006) Early auditory experience generates long-lasting memories that may subserve vocal learning in songbirds. Proc Natl Acad Sci U S A 103:1088–1093.

Rinnert P, Nieder A (2021) Neural code of motor planning and execution during goal-directed movements in crows. J Neurosci 41:4060–4072.

Rinnert P, Kirschhock ME, Nieder A (2019) Neuronal correlates of spatial working memory in the endbrain of crows. Curr Biol 29:2616–2624.e4.

Roberts TF, Mooney R (2013) Motor circuits help encode auditory memories of vocal models used to guide vocal learning. Hear Res 303:48–57.

Roberts TF, Hisey E, Tanaka M, Kearney MG, Chattree G, Yang CF, Shah NM, Mooney R (2017) Identification of a motor-to-auditory pathway important for vocal learning. Nat Neurosci 20:978–986.

Roediger HL 3rd, Butler AC (2011) The critical role of retrieval practice in long-term retention. Trends Cogn Sci 15:20–27.

Sakata JT, Yazaki-Sugiyama Y (2020) Neural circuits underlying vocal learning in songbirds. In: The neuroethology of birdsong (Sakata JT, Woolley SC, Fay RR, Popper AN, eds), pp 29–63. Cham, Switzerland: Springer.

Sanford SE, Lange HS, Maney DL (2010) Topography of estradiol-modulated genomic responses in the songbird auditory forebrain. Dev Neurobiol 70:73–86.

Simpson HB, Vicario DS (1990) Brain pathways for learned and unlearned vocalizations differ in zebra finches. J Neurosci 10:1541–1556.

Tanaka M, Sun FM, Li YL, Mooney R (2018) A mesocortical dopamine circuit enables the cultural transmission of vocal behaviour. Nature 563:117–120.

Tibbetts EA, Dale J (2007) Individual recognition: it is good to be different. Trends Ecol Evol 22:529–537.

Vahaba DM, Hecsh A, Remage-Healey L (2020) Neuroestrogen synthesis modifies neural representations of learned song without altering vocal imitation in developing songbirds. Sci Rep 10:3602.

Vargha-Khadem F, Gadian DG, Watkins KE, Connelly A, Van Paesschen W, Mishkin M (1997) Differential effects of early hippocampal pathology on episodic and semantic memory. Science 277:376–380.

Vates GE, Broome BM, Mello CV, Nottebohm F (1996) Auditory pathways of caudal telencephalon and their relation to the song system of adult male zebra finches. J Comp Neurol 366:613–642.

Vicario DS, Yohay KH (1993) Song-selective auditory input to a forebrain vocal control nucleus in the zebra finch. J Neurobiol 24:488–505.

Weinberger NM (2004) Specific long-term memory traces in primary auditory cortex. Nat Rev Neurosci 5:279–290.

Yanagihara S, Yazaki-Sugiyama Y (2016) Auditory experience-dependent cortical circuit shaping for memory formation in bird song learning. Nat Commun 7:11946.

Yanagihara S, Yazaki-Sugiyama Y (2019) Social interaction with a tutor modulates responsiveness of specific auditory neurons in juvenile zebra finches. Behav Processes 163:32–36.

Yu K, Wood WE, Theunissen FE (2020) High-capacity auditory memory for vocal communication in a social songbird. Sci Adv 6:eabe0440.

Zann RA (1996) The zebra finch: a synthesis of field and laboratory studies. Oxford, UK: Oxford UP.

Zhao W, Garcia-Oscos F, Dinh D, Roberts TF (2019) Inception of memories that guide vocal learning in the songbird. Science 366:83–89.