NEUROSCIENCE

High-capacity auditory memory for vocal communication in a social songbird

K. Yu[1]*, W. E. Wood[1]*, F. E. Theunissen[1,2,3]†

Effective vocal communication often requires the listener to recognize the identity of a vocalizer, and this recognition is dependent on the listener's ability to form auditory memories. We tested the memory capacity of a social songbird, the zebra finch, for vocalizer identities using conditioning experiments and found that male and female zebra finches can remember a large number of vocalizers (mean, 42) based solely on the individual signatures found in their songs and distance calls. These memories were formed within a few trials, were generalized to previously unheard renditions, and were maintained for up to a month. A fast and high-capacity auditory memory for vocalizer identity has not been demonstrated previously in any nonhuman animals and is an important component of vocal communication in social species.

INTRODUCTION

In species with large vocal repertoires and sophisticated social behaviors, learning to interpret vocal signals requires a large capacity memory system. For example, a high-capacity memory for defining sounds of words is needed to process human language semantics (1). Similarly, humans can recognize a large number of individuals based on the sound of their voices as well as linguistic idiosyncrasies (2, 3) and must therefore have formed memories for those unique acoustic features (4). Young humans form these auditory memories rapidly and retain them for long periods in a process called fast mapping (5)—the formation of these auditory memories with few exposures and their maintenance for long periods of time. While the complexity of animal vocal communication pales in comparison with human spoken language (6), auditory memory also plays an important role in the vocal communication of nonhuman social species. In particular, songbirds demonstrate aptitude in several communicative tasks that require auditory memories for vocal signals (7). For example, young male songbirds imitate the song of a tutor that they have stored as an auditory memory (8); some birds can learn the alarm calls from other species to avoid dangerous situations (9) and can even mimic alarm calls of mammals for deceit purposes (10); and territorial birds learn to recognize their neighbors based on their voice, enabling them to identify and react to unfamiliar intruders at the boundaries of their local territory (11).

Individual recognition based on voice also plays a central role for creating and maintaining bonds in social songbird species such as the zebra finch. In the wild, zebra finches are a gregarious and nomadic species, living and traveling in multifamily colonies sometimes comprising more than 100 individuals (12). Zebra finches also mate for life, making strong pair bonds with their partners that are maintained through vocal communication (12, 13). Laboratory studies have shown that their songs have a strong individual signature and can be used to recognize one's mate (14), father (15), and peers (16). Individual recognition by vocalizations is not restricted to song; distance calls (DCs) (17), begging calls (18), and soft contact calls

(19) are also used for individual recognition in juveniles and adults. In previous work, we have shown that all the call types of the zebra finch repertoire are individualized by distinct individual acoustical cues for each call type and that zebra finches could use those cues to discriminate between two vocalizers, irrespective of the call type (20). Given that zebra finches live in large social groups and that vocal communication plays a key role in the creation and maintenance of their social networks, we hypothesized that they might have a high-capacity auditory memory for the acoustic individual signatures found in their calls. We were also interested in investigating whether zebra finches are capable of fast mapping. To answer these questions, we tested the ability of zebra finches to learn to discriminate the identities of unseen vocalizers based on either their song or DC; the song and the DC are the two loud call types in the zebra finch repertoire with strong individual signatures that birds use to recognize and localize each other often without visual contact (20, 21).

RESULTS

We trained male and female zebra finches to recognize several conspecifics by their songs ($n = 19$) or DC ($n = 19$) using a modified go–no go task with food reward (Fig. 1A). To test the birds on a large number of vocalizers, we used a 5-day learning ladder procedure in which subjects began by discriminating one rewarded vocalizer from one nonrewarded vocalizer, while additional vocalizers were added to the test on subsequent days (Fig. 1, B and C). Zebra finches individualize each of their call types, and, although their song and DCs are fairly idiosyncratic and stereotyped, there is also acoustical variability across renditions produced by a single vocalizer (20). Thus, each vocalizer was represented by multiple renditions of its song or DC (Fig. 1B).

The performance of each subject was evaluated on days 4 and 5, after they had had at least 1 day of training on each vocalizer. Overall, task performance was measured using an odds ratio (OR): the odds of interruption for nonrewarded trials (correct responses) divided by the odds of interruption on rewarded trials (incorrect responses). An OR of 1 indicates behavior at chance level, and greater than 1 indicates that the subject successfully distinguished rewarded from nonrewarded trials. Nearly all subjects had ORs significantly greater than 1, indicating that they were successful at this task, both

[1]Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, USA. [2]Department of Psychology, University of California, Berkeley, Berkeley, USA. [3]Department of Integrative Biology, University of California, Berkeley, Berkeley, USA.
*These authors contributed equally to this work.
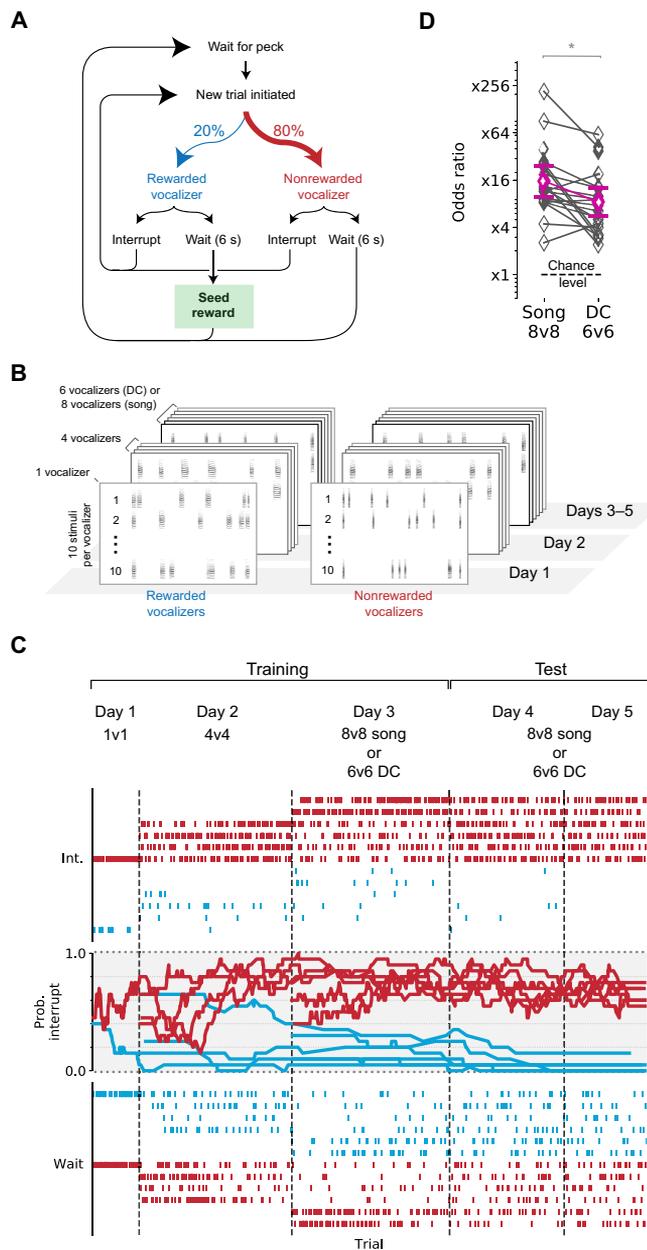†Corresponding author. Email: theunissen@berkeley.edu

Fig. 1. Learning ladder for assessing auditory memory capacity. (**A**) The structure of a single trial. Subjects initiate a trial by pecking a key. A randomly chosen 6-s stimulus file is then played (20% of trials are rewarded, and 80% of trials are nonrewarded). If the stimulus is interrupted by another peck on the same key before the 6-s playback is completed, then a new trial is immediately initiated. If the stimulus is not interrupted and the stimulus is in the rewarded group, then the subject receives 12 s of seed access from a mechanical food hopper. (**B**) The learning ladder procedure gradually introduces new rewarded and nonrewarded vocalizers to the stimulus set each day. Ten stimuli are used for each vocalizer and vocalization type. Each stimulus is, in turn, composed of random sequences of renditions of DCs or songs sampled from our repertoire library for that vocalizer (see also fig. S1 for full-size exemplar spectrograms). (**C**) The lines show the probability of stimulus interruption of individual vocalizers by a single subject in 20 trial bins (blue, rewarded; red, nonrewarded). Tick marks above the plot indicate interrupted trials, and those below the plot indicate noninterrupted trials. (**D**) Average odds ratio (OR) for song and DC assessed after training, on days 4 and 5, for all subjects (n = 19). Birds perform better on songs (OR, 15.5; 95% CI, 9.9 to 24.4) than on DC (OR, 8.4; 95% CI, 5.6 to 12.9) (P = 0.004, log-transformed paired t test). Error bars show 2 SEM.

when tested on songs (19 of 19 subjects) and on DCs (18 of 19 subjects) (P < 0.0026, one-sided Fisher's exact test, Bonferroni corrected; Fig. 1D). There was no difference between males and females on this task as assessed with a mixed effects model, with subject identity as the random effect and call type (DC or song) and subject sex as the fixed effects (fig. S2); the effect of subject sex on the overall log OR was not significant [β = −0.163; 95% confidence interval (CI), −1.012 to 0.687; P = 0.707], and neither was the interaction between subject sex and call type (β = −0.449; 95% CI, −1.315 to 0.416; P = 0.309).

To see whether this performance was driven by memorization of all vocalizers in the test or just recognition of a subset of them, we looked at each subject's performance in detail by evaluating their behavior per individual vocalizer (Fig. 2). We defined the per-vocalizer OR as the ratio of the odds of interrupting a specific vocalizer by the odds of interrupting a random stimulus sampled equally from rewarded and nonrewarded trials. Using this definition, a vocalizer is memorized if the OR is significantly greater than 1 for nonrewarded vocalizers or less than 1 for rewarded vocalizers. We found that 2 of the 19 subjects were able to memorize the entire set of 16 vocalizers from their songs (12 of 19 learned at least half) and 4 of the 19 subjects were able to memorize the entire set of 12 vocalizers from DCs (15 of 19 learned at least half).

To assess the limits of the auditory memory capacity in these songbirds, for four subjects, we intermixed and doubled the size of the two stimulus sets (song and DCs) in the same session. This resulted in a set of DCs from 24 vocalizers and songs from 32 vocalizers for a total of 56 distinct vocalizers. On the first week after completing the two initial learning ladders and testing (song and DC), subjects were trained on the larger song repertoire (16v16) and DC repertoire (12v12) for 3 days each, thus doubling the total number of vocalizers in 6 days. The following week, subjects were given a single day testing session in which previously learned songs and DCs were intermixed for the first time, with only two vocalizers for each rewarding condition and call type. Under this mixed call type condition, subjects continued to self-initiate trials and interrupt the stimuli at rates seen in previous weeks. We then increased the stimulus set to all vocalizers learned thus far (32 vocalizers on song and 24 vocalizers on DC) and evaluated performance on the next 4 days. The results from these four subjects demonstrated that 40, 52, 30, and 47 (mean, 42) vocalizers could be distinguished successfully.

To assess how quickly stimuli were learned, we generated learning curves showing the interruption probability versus the number of informative trials seen, where an "informative trial" is a trial in which the subject did not interrupt the stimulus, giving the bird an opportunity to learn the reward association (interrupted trials do not give the subject new information about whether the stimulus is rewarded or not) (fig. S4). For both songs and DCs, the probability of interrupting rewarded and nonrewarded stimuli is indistinguishable when no informative trials have been seen (intercepts in Fig. 3, A and B), as one would expect. However, the interruption probabilities for rewarded and nonrewarded vocalizers begin to diverge after only a few informative trials, demonstrating very rapid learning of vocalizer's identity (Fig. 3, A and B). There is a significant effect of call type on the rate of this divergence (β = 0.155; 95% CI, 0.086 to 0.222; P < 0.001, mixed effects model), suggesting that songs may be learned more quickly and with fewer examples (Fig. 3, C and D, and fig. S5). One can also notice that the default "baseline" interruption rates differed between songs and DCs when no informative trials have been seen [song baseline, 0.08 ± 0.01 (2 SEM); DC baseline,
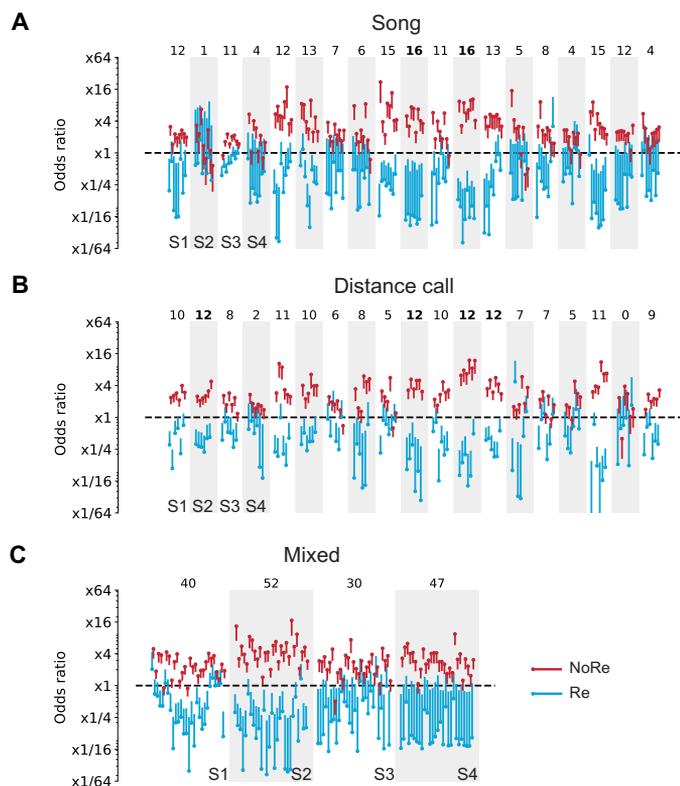
A

B

C

Fig. 2. Memory capacity for vocalizer identity for all subjects. Discrimination performance per vocalizer and subject ($n = 19$) for songs (**A**), DCs (**B**), and both songs and DCs ($n = 4$) (**C**). The mixed condition (C) was performed by four subjects who were additionally tested with a total of 56 vocalizers: 24 vocalizers of DCs and 32 vocalizers of songs. For each subject (white/gray plot background), the dots indicate the OR of interrupting a given vocalizer. Red dots correspond to nonrewarded vocalizers (NoRe) and blue dots to rewarded vocalizers (Re). The number of vocalizers that are discriminated significantly above chance ($P < 0.05$, controlling for false discovery rate using Benjamin-Hochberg procedure) are indicated above each subject's plot (maximum number of vocalizers are 12 for DCs, 16 for songs, and 56 for the mixed condition). Note that the order of the dots on the x axis is random and that the rewarded and nonrewarded vocalizers are not paired. Error bars correspond to the one-sided 95% CI (Fisher's exact test). OR of 1 corresponds to chance. Error bars for nonrewarded stimuli are generally smaller because they are played more frequently. The same data are shown in terms of probabilities in fig. S3.

0.16 ± 0.02; mixed effect models, $P < 0.001$]. The difference in the baseline interruption rates or in the learning rates between male and female subjects was not significant (mixed effects models, $P = 0.563$).

As mentioned above, to encourage subjects to use the individual signature and not a particular acoustical feature present in a given rendition, a vocalizer is represented by randomly chosen call renditions. If subjects are identifying the vocalizer and not memorizing the individual recordings, then they should be able to correctly predict to which reward contingency a novel rendition belongs when they have already heard and learned some of the renditions of a vocalizer. Birds are at chance levels for the first few renditions they hear but begin to correctly categorize previously unheard renditions after exposure to other renditions from the same vocalizer (Fig. 4, A and B); post hoc analysis of the order in which renditions were first presented to subjects reveals that the interruption probability of unseen nonrewarded stimuli increases with the rendition presen-

tation order ($R^2_{adj,song} = 0.90$ and $R^2_{adj,DC} = 0.81$). In the same vein, the interruption probability of rewarded stimuli decreases with the rendition presentation order for song ($R^2_{adj,song} = 0.71$), but the same decrease was not apparent for DC ($R^2_{adj,DC} = 0.00$). The slopes are steeper for the nonrewarded renditions because nonrewarded stimuli are being presented four times more frequently than rewarded stimuli; thus, they are also learned faster. Thus, birds are learning to identify the identity of the vocalizers and do not just memorize the individual sound files.

To test whether these memories are stable over longer times and without any additional reinforcement, we retested two subjects on the largest stimulus set (32 songs and 24 DCs intermixed) after a month during which they were not exposed to any of the vocalizations from the test. While their overall performance slightly decreased from optimal performance during the initial test as measured by the change in log OR [0.12 ± 0.18 (2 SEM) in subject 1 and −0.73 ± 0.23 in subject 2], the overall ORs and OR per vocalizer were still well above chance ($P < 0.001$), indicating that reward associations were retained after a month. To validate that these responses were remembered and not rapidly relearned, we examined the interruption rates for the first informative trials after 1 month and compared them to the rates found for the first informative trials during initial learning (Fig. 4, C and D). These results indicate that these memories for rewarded and nonrewarded vocalizers are stable and can be recalled a month after learning. This is particularly remarkable given that these memories were acquired rapidly and were only reinforced for a short time.

DISCUSSION
Zebra finches have exceptional auditory memory abilities for the individual signature found in their communication calls. We found that they are able to quickly learn to recognize the identity of up to ~40 vocalizers and to maintain these auditory memories for a long period of time. The recognition of vocalizers is a nontrivial task since it requires the extraction of the individual signature present in each call while ignoring the variability across call renditions. Thus, these are not auditory memories for specific sounds but for the information bearing invariant features constituting the individual signature of the vocalizer (20). We showed that zebra finches can learn and memorize this individual signature with a very small number of exposures (<5), can simultaneously remember a large number of these vocalizers, and are able to use these memories to classify call renditions that they have not heard before (generalization).

The memory capacity in zebra finches for recognizing individuals from their vocalizations is large and might exceed the limits that could be tested with our experimental design. We found that 16 vocalizers based on song and 12 vocalizers based on DC could be regularly discriminated by our subjects. When subjects were tested on as many vocalizers as could be practically tested in a single session, birds were able to discriminate up to 52 distinct vocalizers. The capacity of this auditory memory is similar to other forms of avian memory that have been well quantified, such as spatial memories in food-caching birds (22) or visual memories in pigeons (23). Auditory memories for object labels have also been shown in parrots (24) and in some mammals (25), including the exceptional example of Rico, the border collie, who could correctly fetch ~200 distinct objects on vocal commands (26). We also found that birds make an efficient use of informative trials during their very rapid learning, as they are
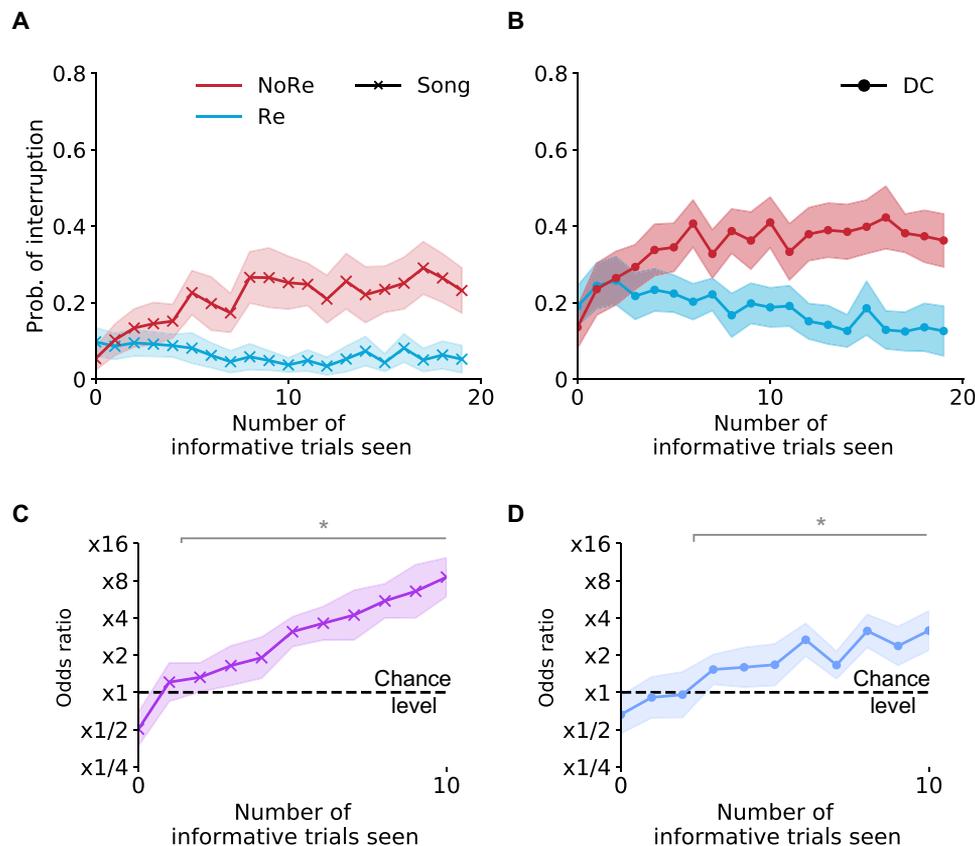
Fig. 3. Speed of memory acquisition. (**A** and **B**) Learning rates are analyzed by plotting the behavioral response (probability of interruption) as a function of informative trials (see Results section) for rewarded (blue) and nonrewarded vocalizers (red). (**C** and **D**) The separation between the red and blue curves in A and B quantifies the learning and is shown in C and D as an OR of odds for nonrewarded divided by the odds of rewarded as in Fig. 1D [(C), song; and (D), DC]. Shaded regions show 2 SEM. Asterisks indicate region where OR was significantly greater than 0 (*n* = 19, *P* < 0.05, false discovery correction).

able to memorize the individual signature of a vocalizer after only a few examples (<10). This fast mapping for communicative vocal signals has only been shown in humans and dogs and is thought to be a key cognitive ability for language learning (*5*, *26*). Last, this memory was long lasting; birds could still remember which vocalizers were assigned to reward versus nonrewarded groups after 1 month without any reinforcement. While previous experiments had shown that song exposure in zebra finches improves auditory recognition, suggestive of a capacity for long-term auditory memories for con-specific vocalizations (*27*), this is the first study that quantifies the auditory memory capacity in a songbird for individual signature and demonstrates its remarkable performance. Just as in humans, we postulate that birds use an abstract neural representation of these auditory objects to facilitate both working memory manipulation and long-term memory storage (*28*).

Since most songbirds are also vocal imitators, one might postulate that the memory mechanisms needed for the song imitation behavior overlap with ones that are needed for individual recognition. The auditory memories could be stored as learned motor programs (*29*), and the high-level abstract representation could then be a motor code. There are many problems with such a motor theory of perception in songbirds: Individual recognition based on vocalizations is present for calls that are not learned (*20*); it is equally similar in male and female zebra finches, while only male zebra finches learn to sing; and male zebra finches learn a single song, but,

as we have shown, they can remember the individual signature of songs and calls from a much larger number of vocalizers. Therefore, although the motor song nuclei might play a role, we and others (*30*) postulate that a separate neural mechanism representing high-level auditory features is involved in the formation and use of memories for all auditory objects that are relevant for vocal communication. The second order avian auditory pallial areas NCM (nidopallium caudal medial) and CM (caudal mesopallium) are good candidates for the locus of such an engram. NCM neurons show neural correlates of memories for the tutor song before vocal learning (*31*), and CM neurons show neural correlates for categories of natural sounds learned in operant conditioning tasks (*32*, *33*). Experiments that have exploited the stimulus-specific habituation observed in NCM neurons also suggest that this auditory area can exhibit a large-capacity memory for conspecific song (*34*). The identity and the connectivity of neural networks involved for storing and recalling these auditory objects as well as the nature of the neural representation for vocalizations, while an active area of research (*35*–*39*), remain relatively unexplored in the birdsong field (*7*). Just as the neural basis of the song imitation behavior has led to many insights into mechanisms of vocal production and learning (*8*), we predict that future work on the neural basis of these auditory memories and their rapid formation will reveal core knowledge of the neural circuits and computations needed for recognizing learned meaning in vocal sounds, including in human speech.
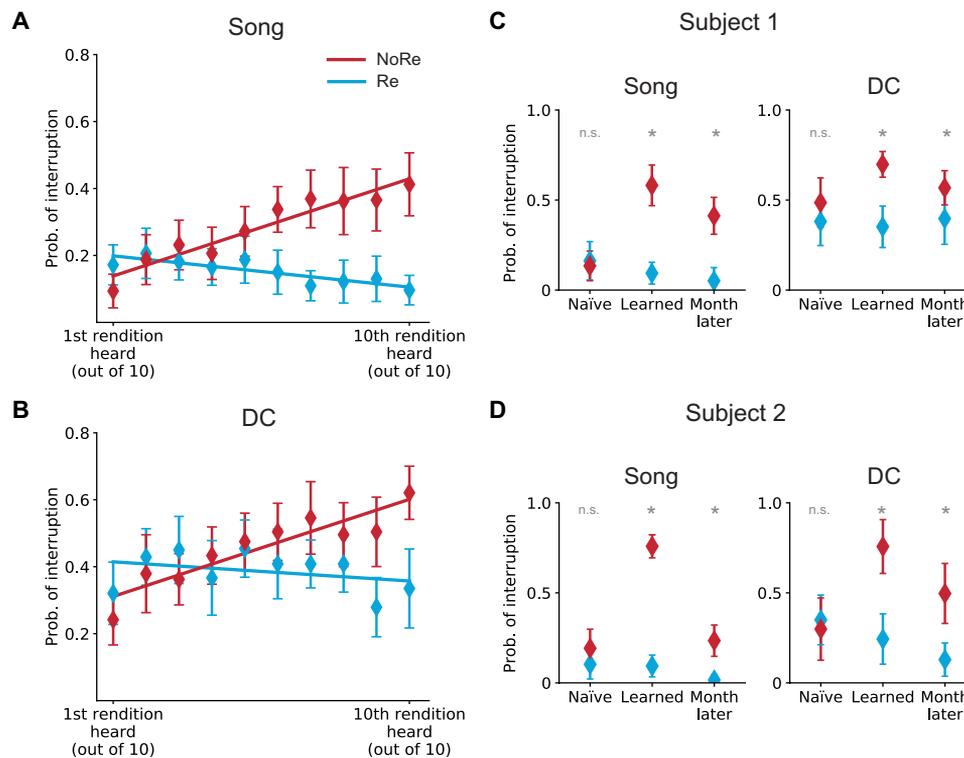
Fig. 4. Generalization and long-term memory. (A and **B**) The plots show the average probability of interruption across all subjects ($n = 19$) for each of the 10 renditions the first time they are heard by the subject; the renditions are ordered on the x axis according to the presentation order. Error bars are 2 SEM. (**C** and **D**) Interruption rates for nonrewarded and rewarded vocalizers in two subjects (S1 and S2 of Fig. 2) during three epochs for songs (left) and DCs (right). The three epochs shown are Naïve (initial exposure to the stimuli), Learned (last two sessions of initial learning ladders), and Month later (1 month after Learned without any reinforcement). The interruption rates to a particular vocalizer are restricted to trials before the second informative trial of that vocalizer during the relevant epoch. Asterisks indicate epochs during which nonrewarded stimuli were interrupted at a significantly higher rate than rewarded stimuli ($P < 0.05$, one-sided t test). Error bars indicate 2 SEM. n.s., not significant.

The fast-learning and exceptional memory for auditory objects in songbirds is a behavioral trait that is essential for vocal communication in social species. This skill can be added to their well-studied vocal imitation behavior, their ability to learn grammar like rules (40, 41), and their capacity to combine call types to generate complex meaning (42). Individual recognition plays an important role for behaviors in social groups and, in particular, for fission-fusion societies such as those observed in some bird species, including the zebra finch (43), and in mammals such as in the African elephant (44). We suggest that these auditory memories for vocalizers are not only important for mate and kin recognition but also to facilitate group dynamics. Studying vocal communication in gregarious bird species should therefore include the role of higher cognitive functions, such as memory, and take into account the species social dynamics. These vocal and perceptual performances can, in turn, be added to the list of cognitive faculties that have been found in social birds, such as episodic spatial memory (22, 45), social cognition (17, 46), number sense (47), or puzzle solving (48), and that rival the cognitive faculties found in social primates (49, 50).

MATERIALS AND METHODS
Ethics statement
All animal procedures were approved by the Animal Care and Use Committee of the University of California, Berkeley (AUP-2016-09-9157) and were in accordance with the National Institutes of Health guidelines regarding the care and use of animals for experimental procedures.

Testing apparatus and software
The operant conditioning apparatus and our go–no go paradigm had been described in detail in our previous publication (20). Briefly, our operant chamber is composed of one pecking key and one food hopper (Med Associates). Subjects initiate trials by pecking the key, which triggers a 6-s auditory stimulus to be played. Sound levels are calibrated to match natural levels of intensity for each call type when vocalizations are used as stimuli. After 6 s, a food reward is either given (if the stimulus was rewarded) or nothing happens (if the stimulus was nonrewarded). Alternatively, as the sound is played, the bird can terminate a trial and start a new one by pecking the same key. In this case, the initial trial will not result in food whether the stimulus is rewarded or not, and a new trial is immediately initiated. To maximize the rate at which reward is received in a session, the subjects learn to skip stimuli that are recognized as nonrewarded to avoid the full 6-s waiting period and move on to the next trial. Subjects are food restricted with access to water but limited seed in between test sessions 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 so that the birds maintain their weight within 10% of their starting weight. Daily handling of subjects did not seem to affect the birds' motivation or ability to do the task once they became comfortable with the

experiment chamber. Once trained, birds are able to get all of their daily food allowance during the testing period.

The birds learn to use the apparatus during a shaping session that lasts approximately 1 week. During the shaping session, the bird first learns to associate pecking of the key with sounds and food reward and then learn to interrupt nonrewarded sounds. The initial shaping task involves the discrimination of two clearly distinct song stimuli. We have also performed control experiments, clearly showing that apparatus is not providing any extraneous clues that the birds could use to distinguish rewarded from nonrewarded trials (20).

The presentation of the sound stimuli, the detection of key pecks, and the operation of the food hopper were controlled by a Python program. We used a custom branch of the Python-based pyOperant software (https://github.com/theunissenlab/pyoperant), originally developed by J. Kiggins and M. Thielk in T. Gentner's laboratory at University of California San Diego (https://github.com/gentnerlab/pyoperant).

Auditory discrimination experiments

Subjects were tasked with discriminating between a set of rewarded and nonrewarded individuals based on the playback of their vocalizations. By design, 20% of trials are rewarded after the end of the stimulus playback, while 80% of trials are not rewarded so that subjects learn to peck for a new trial (interrupting the current trial) when they recognize a stimulus as nonrewarded.

For each vocalizer, we generated 10 unique stimuli that could be played on each trial so that specific extraneous acoustic features of a particular stimulus file that did not encode the vocalizer identity (e.g., length, intensity, and background noise) could not be used as a reward cue. Each song stimulus file consisted of three randomly selected song bouts of two motifs, each from the same vocalizer, separated by randomly chosen intervals such that the duration of the stimulus file would be exactly 6 s. Most introductory notes (repeated short vocalizations preceding a song bout with sometimes long internote intervals) were removed to avoid great variability in stimulus duration. Similarly, each DC stimulus file consisted of six randomly selected DC renditions from one vocalizer, separated by randomly chosen intervals. The amplitudes of the audio files were normalized within stimuli of the same type, i.e., songs or DCs.

On the first day of the test, a subject is tasked with discriminating between one rewarded vocalizer and one nonrewarded vocalizer. Over this single session of about 8 hours, subjects learned to interrupt nonrewarded trials and to wait on rewarded trials. On subsequent days, additional vocalizers were added to the test (Fig. 1): After the first day of 1 rewarded vocalizer versus 1 nonrewarded vocalizer (1v1), we added stimuli from three more rewarded and three more nonrewarded vocalizers, resulting in four rewarded versus four nonrewarded vocalizers (4v4), again with 10 unique renditions per vocalizer. After the day of 4v4, the birds moved on to 8v8 (for songs) or 6v6 (for DCs). Because subjects do as few as ~200 trials per day and we only play rewarded trials 20% of the time, a single vocalizer may be heard as few as five times per day on average once we reach 8v8. We expected that this would make learning at that stage of the ladder difficult. To aid in learning and allow the birds more opportunities to learn every stimulus, on the first day of 8v8 or 6v6, we played stimuli from the new vocalizers twice as frequently as stimuli from vocalizers previously seen on the 1v1 and 4v4 days. On the last 2 days of 6v6/8v8, the probability was set again to be equal across all vocalizers of the same reward outcome. We used these last 2 days to evaluate

task performance. In a few cases, the 1v1 or 4v4 day was repeated (4 of 19 during 1v1 days, 4 of 19 during 4v4 days) because the subject failed to trigger a sufficiently large number of trials.

Vocalizers were randomly assigned to the rewarded or nonrewarded set. Moreover, we used a balanced procedure where the rewarded and nonrewarded sets were switched for each half of the birds in the experiment. Last, for DCs, male and female vocalizers were also randomly assigned to rewarded and nonrewarded sets. The zebra finch DC is sexually dimorphic (21), and by mixing male and female vocalizers in each set, we forced our subjects to use the individual signature and not the acoustic features characteristic of the sex of the vocalizer.

Subjects

Twenty adult domestic zebra finches (10 males and 10 females) were used as subjects in this study. One female subject was excluded from the song memory test analysis due to errors in stimulus selection. A different female subject was excluded from the DC memory test analysis for the same reason, resulting in $n = 19$ for both the song and DC analysis. Subjects were housed in a colony room (usually 10 to 30 individuals in a large flight cage) at the University of California (UC) Berkeley. Of these 20 subjects, 4 subjects were chosen (randomly) to participate in a second session with the combined and larger stimulus set, and 2 of those 4 birds were chosen in the third session to assess long-term memory.

Song vocalization recordings were from 32 male zebra finches from the Theunissen Lab at UC Berkeley, the Perkel laboratory at the University of Washington, and the Leblois laboratory, Bordeaux (France) Neurocampus. 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 unfamiliar 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 (7 switched, 5 same).

Statistical analyses

Performance on the task overall was quantified as an OR obtained by dividing the odds of interrupting a nonrewarded stimulus by the odds of interrupting a rewarded stimulus. The odds of interrupting a stimulus in a given reward group was calculated by taking all trials of that reward category and computing the probability of interruption.

Table 1. Contingency matrix used to estimate the OR of interruption for nonrewarded vs rewarded vocalizer.

	Interruptions	Waits
Nonrewarded	a	c
Rewarded	b	d

Table 2. Contingency matrix used to estimate the OR of interruption for a particular vocalizer relative to a random vocalizer.

	Interruptions	Waits
Vocalizer	a	c
Random	b	d

For Fig. 1C, this was computed on the trials from the last 2 days of tests (6v6 DCs and 8v8 songs) when all vocalizers were played at equal rates. Performance on songs was compared to performance on DCs with a paired t test over subjects. All ORs and 95% CIs were computed using the Fisher's exact test using the contingency matrix shown in Table 1.

The odds of interruption of the nonrewarded stimulus is $O_{\mathrm{NoRe}} = \frac{a}{c}$; similarly, the odds of interruption of the rewarded stimuli is $O_{\mathrm{Re}} = \frac{b}{d}$. The OR is $\mathrm{OR} = \frac{ad}{bc}$. The Fisher's exact test calculates the probability of obtaining an OR as extreme (equal or greater) by calculating the distribution of all ORs obtained for all possible contingency matrices that have the same marginals as those in the actual data. Zero values in any cell cause the OR to be undefined or go to infinity. To avoid this issue, we used the Haldane-Anscombe correction by adding 0.5 to all cells before computing the OR.

Performance per vocalizer was quantified as an OR obtained by dividing the odds of interrupting a given vocalizer by the odds of interrupting a random vocalizer during the time period of interest (Fig. 2). The odds of interrupting a random vocalizer was computed by sampling equal numbers of rewarded and nonrewarded trials on the last 2 days of the 8v8 song and 6v6 DC ladders (Fig. 2, A and B) or over 5 days of the 28v28 mixed set (Fig. 2C), using the contingency matrix shown in Table 2.

Learning curves (Fig. 3) were computed as a function of informative trials, where an informative trial is defined as a trial in which the subject did not interrupt. The probability of interruption in bin k for a subject vocalizer pair is computed by pooling over all trials after the kth interruption and up to and including the $(k + 1)$th non-interruption of that vocalizer. Interruption rates of 0 were adjusted by replacing them with 0.5 times the mean interruption rate across all vocalizers for the same reward contingency in that informative trial bin. Population mean and SEM were then computed across subjects. Significance in bin k was evaluated using the Bonferroni correction. Learning rate is evaluated as the rate at which the log OR between interruption rates on nonrewarded and rewarded trials increases. The effect of call type (song versus DC) on the learning rate was measured using a mixed effects model, with subject as the random effect and call type and informative trials as the fixed effects, predicting the log OR between nonrewarded and rewarded interruptions.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/6/46/eabe0440/DC1

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. M. Brysbaert, M. Stevens, P. Mandera, E. Keuleers, How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Front. Psychol.* **7**, 1116 (2016).
2. V. Aglieri, R. Watson, C. Pernet, M. Latinus, L. Garrido, P. Belin, The Glasgow Voice Memory Test: Assessing the ability to memorize and recognize unfamiliar voices. *Behav. Res. Methods* **49**, 97–110 (2017).
3. T. K. Perrachione, S. N. Del Tufo, J. D. Gabrieli, Human voice recognition depends on language ability. *Science* **333**, 595 (2011).
4. P. Belin, S. Fecteau, C. Bédard, Thinking the voice: Neural correlates of voice perception. *Trends Cogn. Sci.* **8**, 129–135 (2004).
5. L. Markson, P. Bloom, Evidence against a dedicated system for word learning in children. *Nature* **385**, 813–815 (1997).
6. M. D. Hauser, N. Chomsky, W. T. Fitch, The faculty of language: What is it, who has it, and how did it evolve? *Science* **298**, 1569–1579 (2002).
7. J. E. Elie, F. Theunissen, in *The Neuroethology of Birdsong*, J. T. Sakata, S. C. Woolley, R. R. Fay, A. N. Popper, Eds. (Springer, 2020), vol. 71, chap. 7, pp. 175–209.
8. J. T. Sakata, S. C. Woolley, *The Neuroethology of Birdsong*, R. R. Fay, A. N. Popper, Eds. (Springer Handbook of Auditory Research, Springer, 2020), vol. 71.
9. D. A. Potvin, C. P. Ratnayake, A. N. Radford, R. D. Magrath, Birds learn socially to recognize heterospecific alarm calls by acoustic association. *Curr. Biol.* **28**, 2632–2637.e4 (2018).
10. T. P. Flower, M. Gribble, A. R. Ridley, Deception by flexible alarm mimicry in an African bird. *Science* **344**, 513–516 (2014).
11. D. E. Kroodsma, The effect of large song repertoires on neighbor "recognition" in male song sparrows. *The Condor* **78**, 97–99 (1976).
12. R. Zann, *The Zebra Finch: A Synthesis of Field and Laboratory Studies* (Oxford Univ. Press, Oxford, 1996).
13. J. E. Elie, M. M. Mariette, H. A. Soula, S. C. Griffith, N. Mathevon, C. Vignal, Vocal communication at the nest between mates in wild zebra finches: A private vocal duet? *Anim. Behav.* **80**, 597–605 (2010).
14. D. B. Miller, The acoustic basis of mate recognition by female Zebra finches (*Taeniopygia guttata*). *Anim. Behav.* **27**, 376–380 (1979).
15. D. B. Miller, Long-term recognition of father's song by female zebra finches. *Nature* **280**, 389–391 (1979).
16. M. Honarmand, K. Riebel, M. Naguib, Nutrition and peer group composition in early adolescence: Impacts on male song and female preference in zebra finches. *Anim. Behav.* **107**, 147–158 (2015).
17. C. Vignal, N. Mathevon, S. Mottin, Audience drives male songbird response to partner's voice. *Nature* **430**, 448–451 (2004).
18. S. Ligout, F. Dentressangle, N. Mathevon, C. Vignal, Not for parents only: Begging calls allow nest-mate discrimination in juvenile zebra finches. *Ethology* **122**, 193–206 (2016).
19. P. B. D'Amelio, M. Klumb, M. N. Adreani, M. L. Gahr, A. ter Maat, Individual recognition of opposite sex vocalizations in the zebra finch. *Sci. Rep.* **7**, 5579 (2017).
20. J. E. Elie, F. E. Theunissen, Zebra finches identify individuals using vocal signatures unique to each call type. *Nat. Commun.* **9**, 4026 (2018).
21. J. E. Elie, F. E. Theunissen, 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 (2016).
22. R. P. Balda, A. C. Kamil, Long-term spatial memory in clark's nutcracker, *Nucifraga columbiana. Anim. Behav.* **44**, 761–769 (1992).
23. R. G. Cook, D. G. Levison, S. R. Gillett, A. P. Blaisdell, Capacity and limits of associative memory in pigeons. *Psychon. Bull. Rev.* **12**, 350–358 (2005).
24. I. M. Pepperberg, Functional vocalizations by an African grey parrot (Psittacus-erithacus). *Z. Tierpsychol.* **55**, 139–160 (1981).
25. L. M. Herman, D. G. Richards, J. P. Wolz, Comprehension of sentences by bottlenosed dolphins. *Cognition* **16**, 129–219 (1984).
26. J. Kaminski, J. Call, J. Fischer, Word learning in a domestic dog: Evidence for "fast mapping". *Science* **304**, 1682–1683 (2004).
27. R. F. Braaten, M. Petzoldt, A. K. Cybenko, Recognition memory for conspecific and heterospecific song in juvenile zebra finches, *Taeniopygia guttata. Anim. Behav.* **73**, 403–413 (2007).
28. S. Joseph, S. Kumar, M. Husain, T. D. Griffiths, Auditory working memory for objects vs. features. *Front. Neurosci.* **9**, 13 (2015).
29. H. Williams, F. Nottebohm, Auditory responses in avian vocal motor neurons: A motor theory for song perception in birds. *Science* **229**, 279–282 (1985).
30. S. M. H. Gobes, J. J. Bolhuis, Birdsong memory: A neural dissociation between song recognition and production. *Curr. Biol.* **17**, 789–793 (2007).
31. S. Yanagihara, Y. Yazaki-Sugiyama, Auditory experience-dependent cortical circuit shaping for memory formation in juvenile zebra finches. *Nat. Commun.* **7**, 11946 (2016).
32. T. Q. Gentner, D. Margoliash, Neuronal populations and single cells representing learned auditory objects. *Nature* **424**, 669–674 (2003).
33. J. M. Jeanne, J. V. Thompson, T. O. Sharpee, T. Q. Gentner, Emergence of learned categorical representations within an auditory forebrain circuit. *J. Neurosci.* **31**, 2595–2606 (2011).
34. S. J. Chew, D. S. Vicario, F. Nottebohm, A large-capacity memory system that recognizes the calls and songs of individual birds. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1950–1955 (1996).
35. J. M. Jeanne, T. O. Sharpee, T. Q. Gentner, Associative learning enhances population coding by inverting interneuronal correlation patterns. *Neuron* **78**, 352–363 (2013).
36. J. E. Elie, F. E. Theunissen, Meaning in the avian auditory cortex: Neural representation of communication calls. *Eur. J. Neurosci.* **41**, 546–567 (2015).
37. A. S. Kozlov, T. Q. Gentner, Central auditory neurons have composite receptive fields. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 1441–1446 (2016).
38. J. E. Elie, F. E. Theunissen, Invariant neural responses for sensory categories revealed by the time-varying information for communication calls. *PLOS Comput. Biol.* **15**, e1006698 (2019).
39. J. M. Moore, S. M. N. Woolley, Emergent tuning for learned vocalizations in auditory cortex. *Nat. Neurosci.* **22**, 1469–1476 (2019).
40. T. Q. Gentner, K. M. Fenn, D. Margoliash, H. C. Nusbaum, Recursive syntactic pattern learning by songbirds. *Nature* **440**, 1204–1207 (2006).
41. C. ten Cate, The comparative study of grammar learning mechanisms: Birds as models. *Curr. Opin. Behav. Sci.* **21**, 13–18 (2018).

42. T. N. Suzuki, D. Wheatcroft, M. Griesser, Call combinations in birds and the evolution of compositional syntax. *PLOS Biol.* **16**, e2006532 (2018).

43. M. J. Silk, D. P. Croft, T. Tregenza, S. Bearhop, The importance of fission-fusion social group dynamics in birds. *Ibis* **156**, 701–715 (2014).

44. K. McComb, C. Moss, S. Sayialel, L. Baker, Unusually extensive networks of vocal recognition in African elephants. *Anim. Behav.* **59**, 1103–1109 (2000).

45. N. S. Clayton, A. Dickinson, Scrub jays (*Aphelocoma coerulescens*) remember the relative time of caching as well as the location and content of their caches. *J. Comp. Psychol.* **113**, 403–416 (1999).

46. N. J. Emery, J. M. Dally, N. S. Clayton, Western scrub-jays (*Aphelocoma californica*) use cognitive strategies to protect their caches from thieving conspecifics. *Anim. Cogn.* **7**, 37–43 (2004).

47. A. Nieder, Evolution of cognitive and neural solutions enabling numerosity judgements: Lessons from primates and corvids. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **373**, 20160514 (2018).

48. B. Heinrich, T. Bugnyar, Testing problem solving in ravens: String-pulling to reach food. *Ethology* **111**, 962–976 (2005).

49. N. J. Emery, N. S. Clayton, The mentality of crows: Convergent evolution of intelligence in corvids and apes. *Science* **306**, 1903–1907 (2004).

50. N. J. Emery, Cognitive ornithology: The evolution of avian intelligence. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361**, 23–43 (2006).

Citation: K. Yu, W. E. Wood, F. E. Theunissen, High-capacity auditory memory for vocal communication in a social songbird. *Sci. Adv.* **6**, eabe0440 (2020).

Science Advances

High-capacity auditory memory for vocal communication in a social songbird

K. Yu, W. E. Wood and F. E. Theunissen