

Auditory cortical error signals retune during songbird courtship

Caleb Jones, Jesse H. Goldberg 

Department of Neurobiology and Behavior, Cornell University, Ithaca, NY 14853, USA • Department of Biomedical Engineering, Cornell University, Ithaca, NY 14853, USA

Reviewed Preprint

Published from the original preprint after peer review and assessment by eLife.

[About eLife's process](#)

Reviewed preprint posted

November 1, 2023 (this version)

Sent for peer review

August 22, 2023

Posted to bioRxiv

August 8, 2023

 https://en.wikipedia.org/wiki/Open_access Copyright information

Abstract

Auditory feedback is important for vocal learning and control, but it remains unclear how the presence of an audience affects neural representations of self-produced sounds. Here we recorded neural activity in a primary auditory cortical area, Field L, in zebra finches practicing singing alone and directing courtship songs to females. We first discovered that many Field L neurons changed their singing-related discharge patterns during courtship singing, even though the auditory feedback from the bird's own song was similar. We next used syllable-targeted distorted auditory feedback (DAF) to test how auditory error signals depend on courtship context. Though past work showed that dopamine neurons uniformly reduce error signaling during courtship, Field L neurons exhibited heterogeneous error signal re-tuning in the presence of the female. Thus, single neurons in a primary sensory area process feedback from self-produced actions differently during practice and performance.

eLife assessment

This study reports that neural activity in the auditory cortex (field L) of singing male songbirds can be modulated by social context. These potentially **important** findings indicate that the presence of a female conspecific alters the response of auditory cortical neurons to the male bird's own song and to perturbations of auditory feedback that the bird has been trained to expect. While they extend recent work showing that the activity of dopaminergic neurons in songbirds is also affected by an audience, the evidence presented is **incomplete** since it is unclear how much of the apparent modulation of cortical neurons may be due to other factors, such as changes in the recorded neurons or their properties over time, which will require additional analyses to work out.

Introduction

Motivational drive affects attention and associated sensory responses to external events (Allen et al., 2019 ; Aton, 2013 ; Hindmarsh Sten et al., 2021 ). Hungry or thirsty animals are more alert to food or water cues, which evoke larger brainwide neural responses (Allen et al., 2019 ; Burgess et al., 2018 ). Courtship also influences behavioral and neural responses to sensory events (Sakata and Brainard, 2009 , 2008 ; Skals et al., 2005 ; Zhang et al., 2018 ), but it remains unclear how representations of self-produced behaviors change with the presence of an

audience during courtship. Here we recorded auditory cortical neurons in birds singing alone and to females to test how auditory feedback signals depend on courtship-associated changes in motivational state.

Adult male zebra finches practice undirected song alone and also direct courtship song to females (Kao et al., 2008 [↗](#); Zann, 1996 [↗](#)). Both undirected and directed songs combine introductory notes, calls, and a stereotyped sequence of phonologically distinct syllables called motifs (Hyland Bruno and Tchernichovski, 2019 [↗](#); Rajan and Doupe, 2013 [↗](#)). The acoustic structure of undirected and directed motifs is highly similar in adult finches, enabling singing-related neural activity to be precisely aligned and compared across contexts (Kao et al., 2008 [↗](#); Woolley et al., 2014 [↗](#); Zann, 1996 [↗](#)). A first goal of this study was to examine motif-aligned discharge in auditory cortical neurons to test if neural representations depend on courtship state.

A second goal was to test how perceived song errors are represented across performance modes. In past work, perceived errors during undirected singing were experimentally controlled with syllable-targeted distorted auditory feedback (DAF) (Andalman and Fee, 2009 [↗](#); Tumer and Brainard, 2007 [↗](#)). Though DAF does not capture all aspects of natural song learning, it brings song evaluation under experimental control (Ali et al., 2013 [↗](#); Canopoli et al., 2014 [↗](#); Hoffmann et al., 2016 [↗](#)). Birds learn to change the way they sing a syllable to avoid DAF, and several studies support a reinforcement learning framework. First, a dopaminergic (DA) projection from the ventral tegmental area (VTA) to Area X, the striatal nucleus of the song system, is necessary for both natural and DAF-based learning (Duffy et al., 2022 [↗](#); Gadagkar et al., 2016 [↗](#); Hisey et al., 2018 [↗](#); Hoffmann et al., 2016 [↗](#); Xiao et al., 2018a [↗](#); Ali et al., 2013 [↗](#); Harding, 2004 [↗](#); Scharff and Nottebohm, 1991 [↗](#)). Second, Area X projecting DA neurons exhibit reward-prediction error (RPE)-like signals during singing: phasic activations following undistorted renditions and suppressions following distorted ones (Gadagkar et al., 2016 [↗](#)). Third, photoactivation or suppression of DA release in Area X can reinforce or suppress syllable variations (Hisey et al., 2018 [↗](#); Xiao et al., 2018a [↗](#)). The role of RPE-like DA signals in evaluating both reward and song outcomes suggest common principles for birdsong learning and more classic reinforcement learning commonly studied in hungry or thirsty mammals (Chen and Goldberg, 2020 [↗](#)).

Recently we discovered that DA signals retune during courtship. Specifically, DAF-associated DA signals known to evaluate undirected song are uniformly reduced during female-directed singing - as if a bird does not attend to his own mistakes (Roeser et al., 2023 [↗](#)). This discovery raises two possibilities. First, DA error signals might be retuned at the level of VTA, for example by behavioral state-dependent modulation of intrinsic or synaptic excitability of VTA DA neurons (Xiao et al., 2018b [↗](#)). If this is the case, then error responses in upstream inputs to VTA could exhibit similar tuning to auditory feedback during directed and undirected song, but their influence on DA spiking is altered. Another possibility is that brainwide responses to auditory feedback differ when alone versus when courting a female, analogous to how motivational drives such as thirst or hunger retune widespread neural responses to water or food cues (Allen et al., 2019 [↗](#); Burgess et al., 2018 [↗](#)), or how auditory cortical signaling is affected by attention (Ahissar et al., 1992 [↗](#); Hubel et al., 1959 [↗](#)). This idea predicts that auditory representations even in a primary cortical area would change with possible attentional and motivational changes associated with a transition from singing alone to singing to a potential mate.

To test these possibilities we recorded single neuronal activity in Field L in birds singing alone and to females. Field L is a primary auditory cortical area situated at the bottom of a cortical hierarchy that projects into multiple higher auditory areas that, in turn, project to VTA (Bottjer et al., 2000 [↗](#); Foster and Bottjer, 1998 [↗](#); Mandelblat-Cerf et al., 2014 [↗](#); Moore and Woolley, 2019 [↗](#)). Here we report that auditory responses to both bird's own song and DAF-induced errors changed in heterogeneous ways at the transition from undirected to courtship directed singing. Thus a bird's auditory representations of its own song vary across practice and courtship performance modes.

Together with past work showing a re-tuning of premotor signals during courtship (Kao et al., 2008 [↗](#); Singh Alvarado et al., 2021 [↗](#); Woolley et al., 2014 [↗](#)), our discovery that a sensory area retunes is consistent with brainwide changes in neural responsiveness during courtship.

Results

Singing-related discharge in Field L is altered during courtship

Because we were interested in how Field L activity changed during female directed song, we recorded at least 20 song motifs during undirected singing and then presented females to elicit courtship song (n=161 neurons, n=11 birds, Methods). Consistent with past work (Keller and Hahnloser, 2009 [↗](#)), we observed heterogeneous singing-related firing in Field L neurons, ranging from highly temporally precise to variable discharge across motif renditions (Figure 1A-C [↗](#)). To compare neural responses between undirected and directed singing, we focused on motif-aligned activity to compare acoustically similar vocalizations in the undirected and directed states (Figure 1A-F [↗](#)). For each neuron, we analyzed mean firing rates, burst fraction, and the temporal precision of motif-locked discharge depended on courtship state, using analyses previously described (Methods) (Kao et al., 2008 [↗](#); Sakata and Brainard, 2009 [↗](#); Woolley et al., 2014 [↗](#)). Burst fraction was computed as the fraction of spikes occurring in burst events, defined as three (or more) spikes with two (or more) consecutive interspike intervals less than the 25th percentile of the ISI distribution (Methods). The temporal precision of neuronal song locked firing was computed as the intermotif correlation coefficient (IMCC, Methods) (Goldberg and Fee, 2010 [↗](#); Olveczky et al., 2005 [↗](#); Kao et al., 2008 [↗](#); Sakata and Brainard, 2009 [↗](#); Woolley et al., 2014 [↗](#)).

At the population level, neither the temporal precision nor mean firing rate of neurons changed with courtship (Figure 1H-I [↗](#)). A small but significant increase in burst fraction was observed (paired t-test, $p=1.16 \times 10^{-4}$, n=161 neurons, Figure 1J [↗](#)). Thus, Field L neurons did not uniformly change their singing-related firing patterns at the transition to courtship singing. But individual neurons could exhibit significant changes in discharge with changes in courtship state, including increases or decreases in their average motif-locked firing rate (n=24 increase; n=22 decrease, $p<0.01$), burst fraction (n=21 increase; n=3 decrease, $p<0.01$), or precision of timing within the motif (n=11 increase; n=6 decrease, $p<0.01$) (Figure 1D-F [↗](#), p values derived from Monte Carlo shuffles by condition, Methods).

Singing-related performance error signals can retune during courtship

Past work showed that Field L neurons can exhibit responses to distorted auditory feedback (DAF) during undirected singing (Keller and Hahnloser, 2009 [↗](#)). To test if error responses exist and/or retune during courtship, we recorded Field L neurons during lone and directed singing as we controlled perceived song quality with syllable-targeted DAF (Andalman and Fee, 2009 [↗](#); Tumer and Brainard, 2007 [↗](#)). For each bird, a specific syllable was probabilistically targeted with a 50 ms song-like sound played through speakers surrounding the bird (Methods) (Chen et al., 2019 [↗](#); Gadagkar et al., 2016 [↗](#); Hamaguchi et al., 2014 [↗](#)). To quantify error responses across the population of neurons and across conditions, we compared the activity between randomly interleaved renditions of distorted and undistorted songs. For each neuron and for each behavioral condition (alone or directed), we first performed the same analysis as in a previous Field L recording study during undirected singing (Keller and Hahnloser, 2009 [↗](#)) (Methods). This analysis identified 58/161 neurons as error responsive in undirected song, and 55/161 during directed singing, and 33/161 exhibiting error responses during both conditions. Visual inspection of error responses in motif-aligned rasters revealed that while some of these neurons unambiguously exhibited robust DAF responses (Figure 2D [↗](#)), many exhibited extremely subtle ones (Figure 2-figure supplement 2A-B [↗](#)).

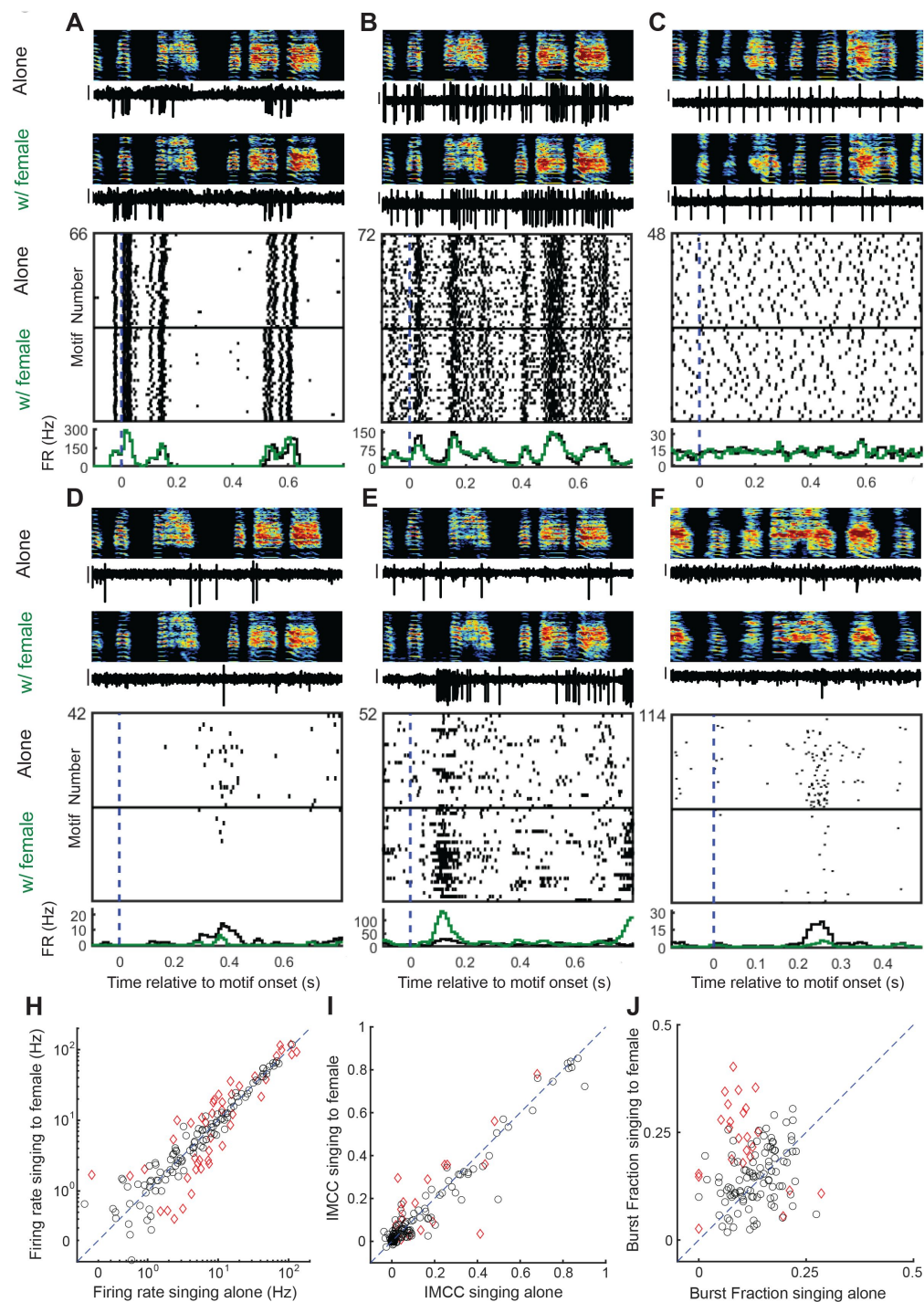


Figure 1.

Singing related discharge can retune during courtship.

(A-C) Three example neurons with stable firing properties across alone and female-directed song conditions. Top to bottom: single trial example spectrograms, spike discharge, corresponding spike raster plots, and rate histograms (aligned to motif onset) from motif renditions singing alone (black) and to the female (green). Black vertical scale bar for spiking activity is 0.2 mV, y-axis limits of spectrograms are 0 to 8 kHz. (D-F) Data plotted as in A-C for three example neurons with significant context-dependent changes in firing. (H-J) Scatter plots of mean firing rates (H), IMCC values (I) and neuronal burst fraction (J) for 161 neurons when singing alone and to the female. Red: neurons with significant change across conditions ($p < 0.01$, Methods).

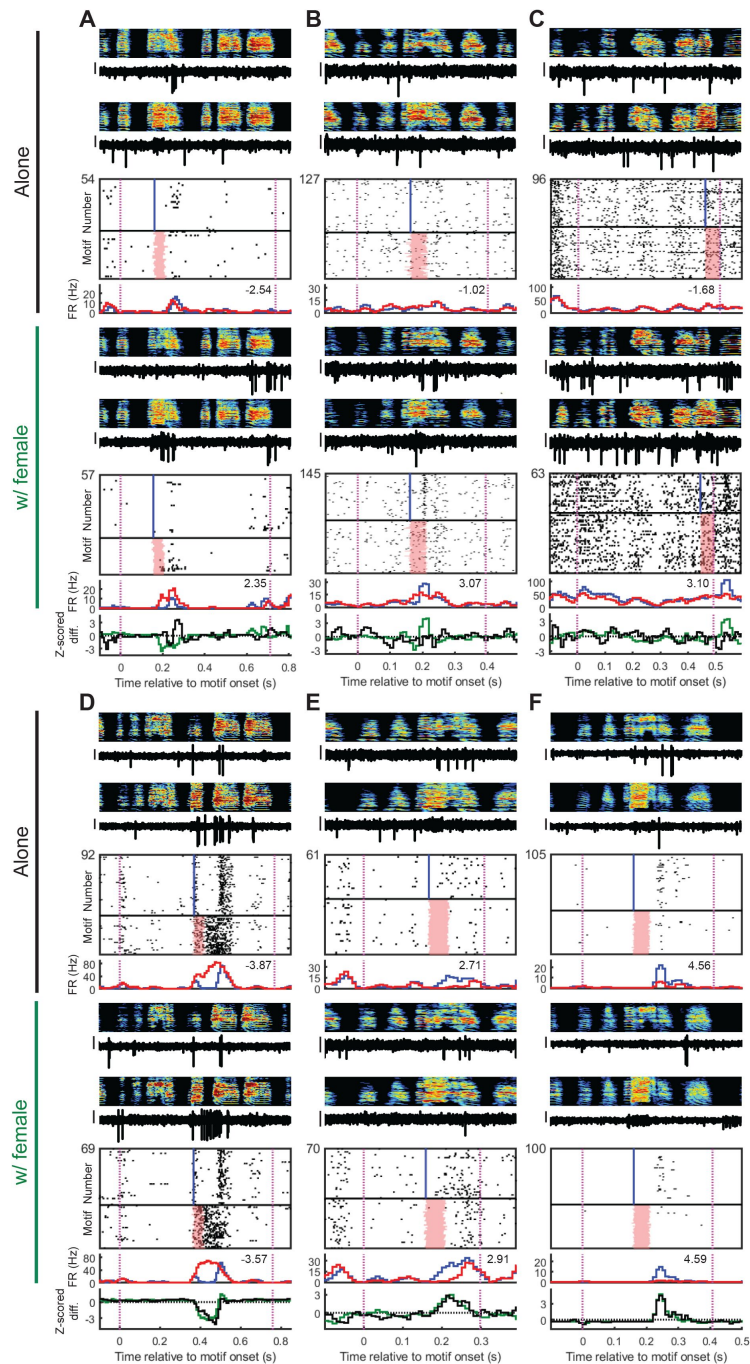


Figure 2.

Error responses in Field L neurons can retune during courtship.

(A-C) Three example neurons with different error responses across courtship conditions. Top to bottom: single trial example spectrograms and spiking activity for undistorted and distorted trials, corresponding raster plots (blue vertical bar denotes feedback target time in undistorted renditions; red shading denotes actual feedback time on distorted renditions; pink vertical dotted line denotes onset and offset of song motif). Corresponding rate histograms for undistorted renditions (blue trace) and distorted renditions (red trace). Below are the same plots, but for songs directed to the female. Bottom: z-scored difference between undistorted and distorted rate histograms for singing alone (black trace) and singing to female (green trace). All data are time-aligned to the onset of the motif. (D-F) Data plotted as in A-C for three neurons with similar error responses across courtship conditions. Error scores for each neuron and condition are enumerated as insets in the histogram. Scale bar for spiking activity is 0.2 mV. Vertical axis limits for spectrograms are 0 to 8 kHz.

Statistical tests defining Field L neurons as error-responsive or not in a binary fashion may not be suitable if the underlying population of error responses exist on a continuum from error non-responsive to responsive. To score each neuron's error response and identify the shape of the error response distributions across the population of neurons, we computed the z-scored difference between target onset-aligned distorted and undistorted rate histograms (target onset defined as the median DAF onset time relative to distorted syllable onset, $n = 161$ neurons in 12 birds), as previously described (Gadagkar et al., 2016). We defined the error response as the average z-scored difference in firing between undistorted and distorted renditions during the 100 ms interval following target onset (Methods). In past recordings from VTA, this analysis yielded a bimodal distribution of z-scored error responses which made classification of error-responsive neurons straightforward. Yet when we plotted the distribution of error responses across the 161 Field L neurons recorded during both undirected and directed singing, we observed unimodal distributions consistent with a continuum of Field L responses (**Figure 2-figure supplement 1**). This analysis suggests that no true threshold exists to unambiguously define a neuron as an error neuron. To visualize a distribution of error scores for parts of the song that were never targeted with DAF, we repeated the same analysis for a time window of the motif preceding the target time (i.e. when the auditory feedback was undistorted across all trials). As expected, we observed significantly lower error scores (**Figure 2-figure supplement 1**, $p < 10^{-10}$, Wilcoxon signed-rank test, Methods). For the purposes of examining courtship-state dependent re-tuning of error signals we conservatively defined 'error neurons' as those with an absolute z-scored difference in firing rate between distorted and undistorted renditions greater than 2.5 in either the directed or undirected conditions (Methods).

To test if courtship state affected these most robust error responses in our dataset, for each error neuron we compared the z-scored error response across undirected and directed singing. Interestingly, though VTAX DA neurons uniformly decreased their error response during courtship singing (Roeser et al., 2023), Field L neurons exhibited heterogenous re-tuning: some neurons exhibited error responses only when singing to the female ($n=22$), others exhibited error signal attenuation during female-directed song ($n=11$), while others were not significantly affected by courtship state ($n=10$, **Figure 2**). Together, these data show that Field L responses to distorted auditory feedback during singing can retune during courtship.

Discussion

By recording from the primary auditory cortex in zebra finches singing alone and to females, we discovered, for the first time to our knowledge, that auditory representations of an animal's own vocalizations change with an audience. These findings extend past work in finches showing that social context affects auditory responses to the calls of conspecifics (Angeloni and Geffen, 2018; Menardy et al., 2014, 2012; Remage-Healey et al., 2010). More broadly, these findings support the idea that auditory cortical activity is modulated not just by acoustic features (Gervain and Geffen, 2019; King et al., 2018) but also diverse phenomena such as attention (Hubel et al., 1959), primary rewards (David et al., 2012), task parameters (Downer et al., 2015; Fritz et al., 2003; Angeloni and Geffen, 2018; Menardy et al., 2014, 2012; Remage-Healey et al., 2010), and even hormone levels (Angeloni and Geffen, 2018; Menardy et al., 2014, 2012; Remage-Healey et al., 2010).

Detecting differences between predicted and actual sensory feedback is a crucial aspect of motor learning (Keller and Mrcic-Flogel, 2018), and auditory cortical neurons in diverse species can signal mismatch errors (Eliades and Wang, 2008; Keller and Hahnloser, 2009; Parras et al., 2021; Ulanovsky et al., 2003). In past recordings from VTA we observed a bimodal distribution of error responses, and the error-responding neurons were the ones that projected to Area X (Gadagkar et al., 2016). Yet in cortex there is uncertainty about the suitability of classifying single neurons by response profile, as task-relevant parameters that can be decoded from

neuronal populations may not be apparent when examining single neuronal representations (Kaufman et al., 2014 [↗](#); Mante et al., 2013 [↗](#); Rigotti et al., 2013 [↗](#); Williams et al., 2018 [↗](#)). We observed a continuum of error responses during both directed and undirected singing which made it difficult to unambiguously define a neuron as being error responsive or not.

Because the main goal of this study was to test if courtship-associated reduction in error signaling, recently observed in VTA DA neurons (Roeser et al., 2023 [↗](#)), resulted from a local process in VTA or reflected a brainwide re-tuning of auditory responsiveness, we imposed an arbitrary threshold of 2.5 in the z-scored error response to test if the most robust error signals in Field L depended on courtship state. Surprisingly, we discovered that Field L neurons could retune at the transition from lone to courtship singing in diverse ways, consistent with a brainwide process that does not fully explain the uniform error signal attenuation observed in VTA. Interestingly, Area X and LMAN neurons uniformly increase their temporal precision and reduce their burstiness during courtship singing (Kao et al., 2008 [↗](#); Roeser et al., 2023 [↗](#); Sakata and Brainard, 2008 [↗](#); Singh Alvarado et al., 2021 [↗](#); Woolley et al., 2014 [↗](#)). In contrast, Field L neurons could exhibit increases or decreases in burstiness, temporal precision, or mean rate with courtship.

An open question is how Field L receives information about whether or not a female is present, and how this information influences neural activity. Several neuromodulatory systems with possible information about courtship state project to songbird auditory forebrain, including acetylcholine (Shea and Margoliash, 2010 [↗](#)), serotonin (Yip et al., 2020 [↗](#)), and norepinephrine (Cardin and Schmidt, 2004 [↗](#)). Estrogens can also rapidly modulate auditory firing (Remage-Healey et al., 2010 [↗](#); Scarpa et al., 2022 [↗](#)). Yet how the courtship state may be orchestrated across multiple brain regions remains unclear. One possibility is that hypothalamic nuclei such as the medial preoptic nucleus (MPOA) initiate the courtship state. The MPOA projects to multiple brainstem neuromodulatory areas which in turn project broadly throughout the forebrain, including the song system and auditory system (Riters and Alger, 2004 [↗](#)) (Ben-Tov et al., 2023 [↗](#); Castelino and Ball, 2005 [↗](#); Singh Alvarado et al., 2021 [↗](#)). It will be interesting in future studies to examine the neural signals propagating through these pathways at transitions into and out of the courtship state.

Materials and Methods

Subjects

Sixteen adult male zebra finches (>90 days post hatch) were the subjects of this study. Animal care and experiments were carried out in accordance with NIH guidelines and were approved by the Cornell Institutional Animal Care and Use Committee.

Surgery and awake-behaving electrophysiology

For chronic neural recordings, subjects were anesthetized with isoflurane inhalation and mounted on a stereotaxic instrument for probe implantation. All probes were 16-channel moveable electrode bundles (Innovative Neurophysiology). Probes were implanted 1.5–2.0 mm anterior and 1.5–2.0 mm lateral of the bifurcation of the mid-sagittal sinus to target Field L (Keller & Hahnloser 2009 [↗](#)) at a head angle of 80 degrees (measured as the angle from the tip of the beak to the center of the ear bars relative to the horizontal plane). The end of the cannula was implanted 1.5 mm ventral to the surface of the brain. Birds were then placed alone in a sound isolation chamber (12 hour light/dark cycle) with ad libitum food and water and allowed 1 day to recover post op before being subjected to the distorted auditory feedback (DAF) protocol. 2–3 days were allotted for habituation to DAF and to ensure the bird began to spontaneously sing a sufficient number of motifs in social isolation before neural recordings. DAF was implemented with a custom LabView acquisition program that analyzed song syllables in real-time and delivered syllable-targeted feedback. DAF (50 ms broadband noise bandpass filtered at 1.5–8 kHz to match frequency range of

zebra finch song) was played over speakers in the recording chamber on top of a specific target syllable randomly on 50% of motif renditions. Experiments were carried out in the male's home cage, which was inside a sound isolation chamber. When the homecage lights came on each day, recording began and the male was left alone to sing at least 40 undirected song motifs. Female directed motifs were then recorded by presenting a female in the chamber in ~10 minute intervals throughout the day until at least 40 directed song motifs were elicited (ref andreas paper). Electrode placement was verified at the end of the experiment with small electrolytic lesions, histology, and dark field imaging. 11 of the 16 implanted birds yielded single unit recordings and sang sufficient motifs for the experiment. Many channels on the probes recorded multi-unit activity, which were taken note of but not analyzed in this study.

Neural Recording and Analysis

Neural signals were acquired with the Intan RHD recording controller and 16-channel Intan headstages that directly interfaced with the moveable bundles. Sampling rate was set to 20kHz and recording was manually controlled with the Intan recording software, where a 60 Hz notch filter was applied and spiking activity of single units could be visualized in real time. Audio data and a real-time digital copy of the DAF signal were simultaneously recorded with neural data in the recording controller such that all data could be easily time aligned. A custom MATLAB GUI was used for visualizing song and neural data, and for spike sorting. Neural recordings were bandpass filtered between 0.4 kHz and 6-8 kHz and single unit spiking activity was manually sorted as previously described (Goldberg and Fee, 2010). Motif aligned spiking activity was time-warped to the median duration of undirected or directed motifs. Firing rate (FR) histograms were computed by binning spiking events in 10 ms windows and smoothing with a 3-bin moving average. To calculate the significance of the FR changes across behavioral contexts, we randomly assigned spike trains from each song motif trial as undirected or directed groups while conserving trial numbers, then calculated FR values for the randomized dataset, as previously described (Goldberg and Fee, 2010). This shuffling without replacement was repeated 10,000 times for each neuron, yielding 10,000 new changes in FR values. Measured changes in firing rate that were greater than the 99th percentile of the shuffled distribution were considered significant, as in previous studies (Sakata and Brainard, 2008). This procedure was repeated for IMCC and burst fraction. A p value less than 0.01 was considered significant to account for multiple comparisons. To quantify the degree to which neuronal firing was time-locked to song, the intermotif correlation coefficient (IMCC) was calculated as described previously (Chen et al., 2019; Kao et al., 2008; Olveczky et al., 2005). To compute IMCC, motif-aligned FR was mean-subtracted and smoothed with a Gaussian kernel of 20 ms SD, resulting in a rate vector, $\mathbf{r}_i$, for each motif. IMCC was defined as the mean of all pairwise correlation coefficients between $\mathbf{r}_i$ as follows:

$$IMCC = \frac{1}{N_{pairs}} \sum_{j>i}^{N_{pairs}} CC_{ij}$$

$$CC_{ij} = \frac{\mathbf{r}_i \cdot \mathbf{r}_j}{\sqrt{r_i^2 r_j^2}}$$

We defined bursts as events containing three or more spikes with consecutive interspike intervals less than the 25th percentile of the interspike interval distributions during singing. Burst fraction was calculated as the fraction of spikes during song motifs that occurred during bursts.

We quantified the error responsiveness for each neuron in two ways, following approaches previously described (Chen et al., 2019; Gadagkar et al., 2016; Keller and Hahnloser, 2009). Spike counts in 30 ms time windows shifted in 5 ms steps up to 50 ms after DAF offset were generated for undistorted and distorted trials. A WRS test assessed the significance of each 30 ms window, and only neurons with 2 or more consecutively significant windows were considered a significant error response. Neurons with two or more subsequent windows with $P < 0.05$ were

considered significant. (Keller and Hahnloser, 2009 [↗](#)). We also calculated the z-scored difference between smoothed firing rates in undistorted and distorted trials for the time window 100 ms after feedback onset, and defined the error score as the average z-scored difference across the three bins centered around the absolute maximum difference during the response window. To test if error responses were attributable to DAF, we compared absolute error score distributions derived from the same analysis in the 100 ms window preceding target onset in the motif and therefore not associated with DAF. The pre-DAF error score distribution was significantly less in both conditions: a paired Wilcoxon signed-rank test indicated significantly lower error scores pre-DAF than post-DAF during undirected song ($p < 0.001$, median pre-DAF: 0.82, median post-DAF: 1.39) and directed song ($p < 0.001$, median pre-DAF: 0.81, median post-DAF: 1.41). 3.4% of pre-DAF and 16.5% of post-DAF error scores were greater than 2.5. Given the continuum of post-DAF error scores in both directed and undirected conditions (**Figure 2-figure supplement 1** [↗](#)), we conservatively chose a threshold of 2.5 to define an error response. Neurons with error responses greater than 2.5 in only one condition (undirected versus directed) were considered to have retuned; neurons with error scores greater than 2.5 in both conditions were considered not to have retuned. Our results did not fundamentally change with an even more strict definition for an error response. With a more stringent threshold of 3, 12 neurons exhibit error response only when singing to the female, and 8 neurons exhibit error signal attenuation during female-directed song, and 6 neurons exhibit a stable error response across conditions.

Figure Legends

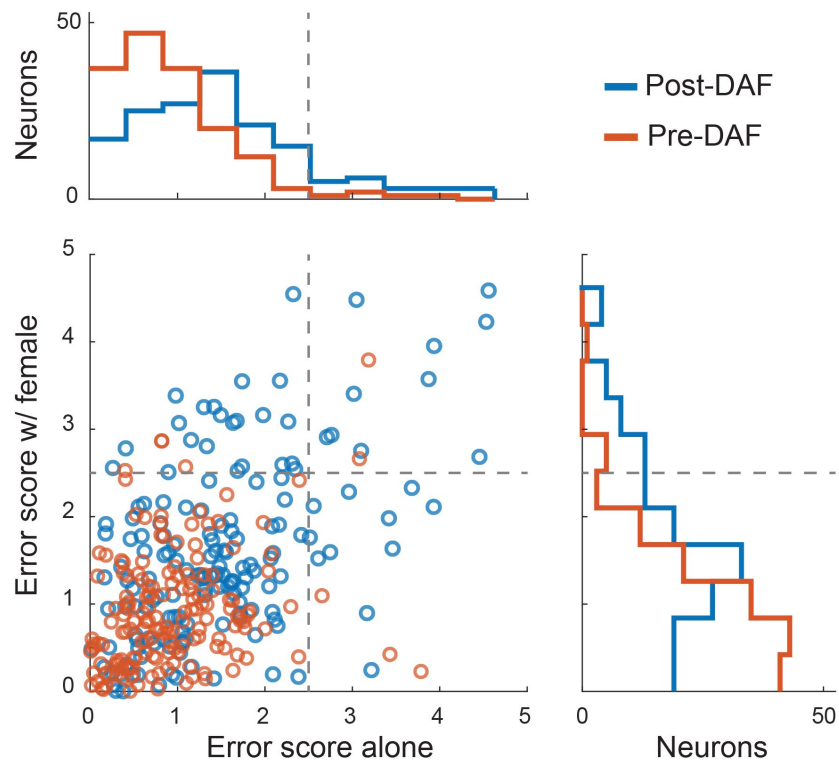


Figure S1-figure supplement 1.

Distributions of absolute error responses in Field L neurons

Scatter plot where each dot represents the absolute z-scored error response of a single neuron during the 100 ms interval after the onset of DAF (blue) and 100 ms before DAF onset (orange) (Methods). Corresponding histograms for each condition are projected along the x and y axes. Gray dotted lines indicate the 2.5 cutoff threshold used to define error neurons.

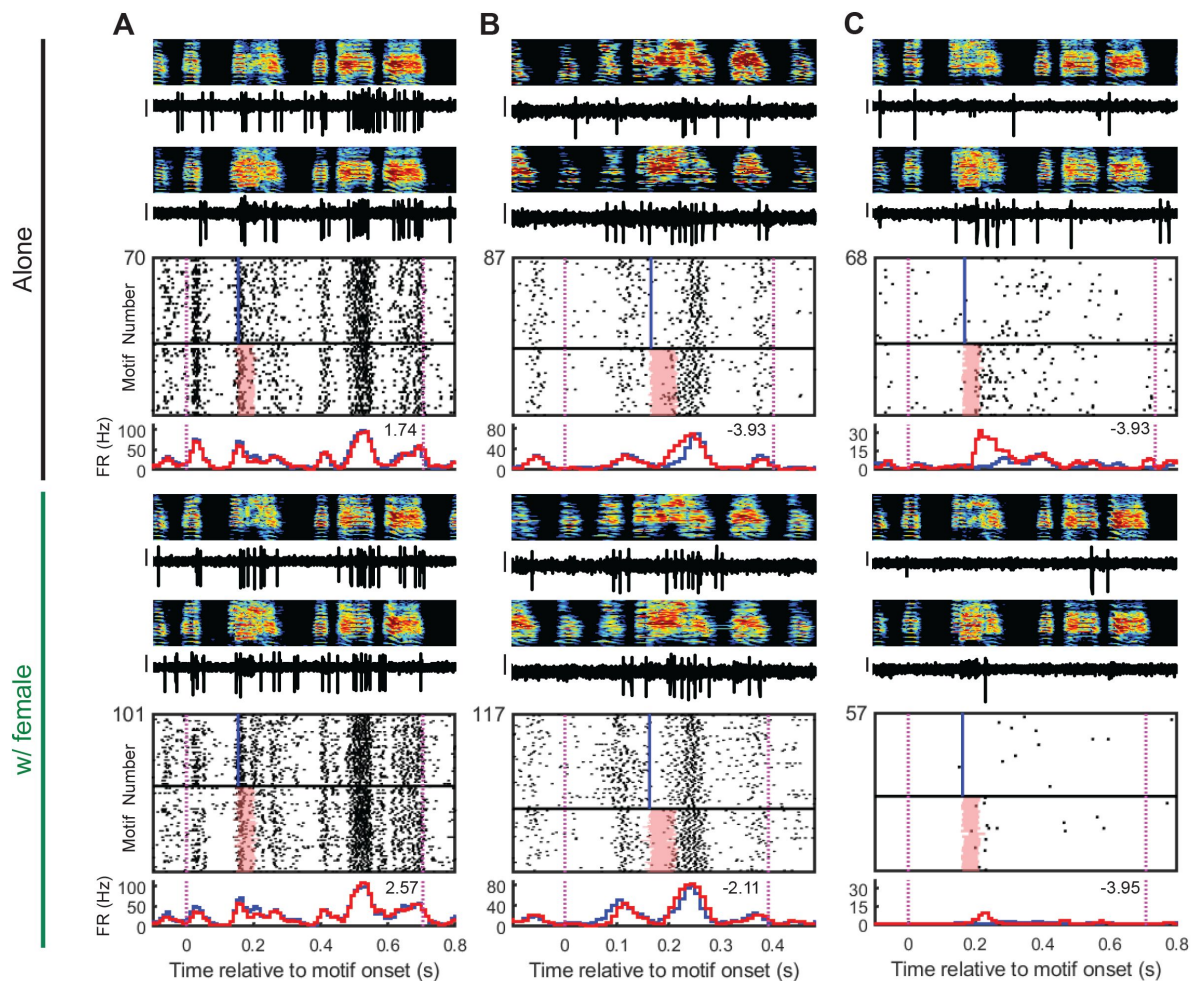


Figure S2-figure supplement 2.

Neurons exemplifying a broad range of error responses in Field L.

(A-C) Top to bottom: example spectrograms, spike discharge, corresponding spike raster plots, and rate histograms for undistorted (blue) and distorted (red) trials aligned to motif onset (blue vertical bar denotes feedback target time in undistorted renditions; red shading denotes DAF; pink vertical dotted line denotes onset and offset of song motif). Error scores for each neuron and condition are enumerated as insets in the histogram. Scale bar for spiking activity is 0.2 mV. Vertical axis limits for spectrograms are 0 to 8 kHz.

References

- Ahissar E, Vaadia E, Ahissar M, Bergman H, Arieli A, Abeles M. (1992) **Dependence of cortical plasticity on correlated activity of single neurons and on behavioral context** *Science* **257**:1412–1415
- Ali F, Otchy TM, Pehlevan C, Fantana AL, Burak Y, Olveczky BP (2013) **The basal ganglia is necessary for learning spectral, but not temporal, features of birdsong** *Neuron* **80**:494–506
- Allen WE, Chen MZ, Pichamoorthy N, Tien RH, Pachitariu M, Luo L, Deisseroth K. (2019) **Thirst regulates motivated behavior through modulation of brainwide neural population dynamics** *Science* **364**
- Andalman AS, Fee MS (2009) **A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors** *Proc Natl Acad Sci U S A* **106**:12518–12523
- Angeloni C, Geffen MN (2018) **Contextual modulation of sound processing in the auditory cortex** *Curr Opin Neurobiol* **49**:8–15
- Aton SJ (2013) **Set and setting: how behavioral state regulates sensory function and plasticity** *Neurobiol Learn Mem* **106**:1–10
- Ben-Tov M, Duarte F, Mooney R. (2023) **A neural hub for holistic courtship displays** *Curr Biol* **33**:1640–1653
- Bottjer SW, Brady JD, Cribbs B. (2000) **Connections of a motor cortical region in zebra finches: relation to pathways for vocal learning** *J Comp Neurol* **420**:244–260
- Burgess CR, Livneh Y, Ramesh RN, Andermann ML (2018) **Gating of visual processing by physiological need** *Curr Opin Neurobiol* **49**:16–23
- Canopoli A, Herbst JA, Hahnloser RH (2014) **A higher sensory brain region is involved in reversing reinforcement-induced vocal changes in a songbird** *J Neurosci* **34**:7018–7026
- Cardin JA, Schmidt MF (2004) **Noradrenergic inputs mediate state dependence of auditory responses in the avian song system** *Journal of Neuroscience* **24**:7745–7753
- Castelino CB, Ball GF (2005) **A role for norepinephrine in the regulation of context-dependent ZENK expression in male zebra finches (*Taeniopygia guttata*)** *Eur J Neurosci* **21**:1962–1972
- Chen R, Goldberg JH (2020) **Actor-critic reinforcement learning in the songbird** *Curr Opin Neurobiol* **65**:1–9
- Chen R, Puzerey PA, Roeser AC, Riccelli TE, Podury A, Maher K, Farhang AR, Goldberg JH (2019) **Songbird Ventral Pallidum Sends Diverse Performance Error Signals to Dopaminergic Midbrain** *Neuron* **103**:266–276
- David SV, Fritz JB, Shamma SA (2012) **David SV, Fritz JB, Shamma SA. 2012. Task reward structure shapes rapid receptive field plasticity in auditory cortex** **109**:2144–2149. *Task reward structure shapes rapid receptive field plasticity in auditory cortex* **109**:2144–2149

- Downer JD, Niwa M, Sutter ML (2015) **Task engagement selectively modulates neural correlations in primary auditory cortex** *J Neurosci* **35**:7565–7574
- Duffy A, Latimer KW, Goldberg JH, Fairhall AL, Gadagkar V. (2022) **Dopamine neurons evaluate natural fluctuations in performance quality** *Cell Rep* **38**
- Eliades SJ, Wang X. (2008) **Neural substrates of vocalization feedback monitoring in primate auditory cortex** *Nature* **453**:1102–1106
- Foster EF, Bottjer SW (1998) **Axonal connections of the high vocal center and surrounding cortical regions in juvenile and adult male zebra finches** *J Comp Neurol* **397**:118–138
- Fritz J, Shamma S, Elhilali M, Klein D. (2003) **Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex** *Nat Neurosci* **6**:1216–1223
- Gadagkar V, Puzerey PA, Chen R, Baird-Daniel E, Farhang AR, Goldberg JH (2016) **Dopamine neurons encode performance error in singing birds** *Science* **354**:1278–1282
- Gervain J, Geffen MN (2019) **Efficient Neural Coding in Auditory and Speech Perception** *Trends Neurosci* **42**:56–65
- Goldberg JH, Fee MS (2010) **Singing-related neural activity distinguishes four classes of putative striatal neurons in the songbird basal ganglia** *J Neurophysiol* **103**:2002–2014
- Hamaguchi K, Tschida KA, Yoon I, Donald BR, Mooney R. (2014) **Auditory synapses to song premotor neurons are gated off during vocalization in zebra finches** *Elife* **3**
- Harding CF (2004) **Brief alteration in dopaminergic function during development causes deficits in adult reproductive behavior** *J Neurobiol* **61**:301–308
- Hindmarsh Sten T, Li R, Otopalik A, Ruta V. (2021) **Sexual arousal gates visual processing during Drosophila courtship** *Nature* **595**:549–553
- Hisey E, Kearney MG, Mooney R. (2018) **A common neural circuit mechanism for internally guided and externally reinforced forms of motor learning** *Nat Neurosci* **1**
- Hoffmann LA, Saravanan V, Wood AN, He L, Sober SJ (2016) **Dopaminergic Contributions to Vocal Learning** *J Neurosci* **36**:2176–2189
- Hubel DH, Henson CO, Rupert A, Galambos R. (1959) **“attention” units in the auditory cortex** *Science* **129**:1279–1280
- Hyland Bruno J, Tchernichovski O. (2019) **Regularities in zebra finch song beyond the repeated motif** *Behav Proc* **163**
- Kao MH, Wright BD, Doupe AJ (2008) **Neurons in a forebrain nucleus required for vocal plasticity rapidly switch between precise firing and variable bursting depending on social context** *J Neurosci* **28**:13232–13247
- Kaufman MT, Churchland MM, Ryu SI, Shenoy KV (2014) **Cortical activity in the null space: permitting preparation without movement** *Nat Neurosci* **17**:440–448
- Keller GB, Hahnloser RHR (2009) **Neural processing of auditory feedback during vocal practice in a songbird** *Nature* **457**:187–190

- Keller GB, Mrsic-Flogel TD (2018) **Predictive Processing: A Canonical Cortical Computation** *Neuron* **100**:424–435
- King AJ, Teki S, Willmore BDB (2018) **Recent advances in understanding the auditory cortex [version 1; peer review: 2 approved]** *F1000Res* **7** <https://doi.org/10.12688/F1000RESEARCH.15580.1>
- Mandelblat-Cerf Y, Las L, Denisenko N, Fee MS (2014) **A role for descending auditory cortical projections in songbird vocal learning** *Elife* **3** <https://doi.org/10.7554/eLife.02152>
- Mante V, Sussillo D, Shenoy KV, Newsome WT (2013) **Context-dependent computation by recurrent dynamics in prefrontal cortex** *Nature* **503**:78–84
- Menardy F, Giret N, Del Negro C. (2014) **The presence of an audience modulates responses to familiar call stimuli in the male zebra finch forebrain** *Eur J Neurosci* **40**:3338–3350
- Menardy F, Touiki K, Dutrieux G, Bozon B, Vignal C, Mathevon N, Del Negro C. (2012) **Social experience affects neuronal responses to male calls in adult female zebra finches** *European Journal of Neuroscience* <https://doi.org/10.1111/j.1460-9568.2012.08047.x>
- Moore JM, Woolley SMN (2019) **Emergent tuning for learned vocalizations in auditory cortex** *Nat Neurosci* **22**:1469–1476
- Olveczky BP, Andalman AS, Fee MS (2005) **Vocal experimentation in the juvenile songbird requires a basal ganglia circuit** *PLoS Biol* **3**
- Parras GG, Casado-Román L, Schröger E, Malmierca MS (2021) **The posterior auditory field is the chief generator of prediction error signals in the auditory cortex** *Neuroimage* **242**
- Rajan R, Doupe AJ (2013) **Behavioral and neural signatures of readiness to initiate a learned motor sequence** *Curr Biol* **23**:87–93
- Remage-Healey L, Coleman MJ, Oyama RK, Schlinger BA (2010) **Brain estrogens rapidly strengthen auditory encoding and guide song preference in a songbird** *Proc Natl Acad Sci U S A* **107**:3852–3857
- Rigotti M, Barak O, Warden MR, Wang XJ, Daw ND, Miller EK, Fusi S. (2013) **The importance of mixed selectivity in complex cognitive tasks** *Nature* **497**:585–590
- Riters LV, Alger SJ (2004) **Neuroanatomical evidence for indirect connections between the medial preoptic nucleus and the song control system: possible neural substrates for sexually motivated song** *Cell Tissue Res* **316**:35–44
- Roeser A, Gadagkar V, Das A, Puzerey PA, Kardon B, Goldberg JH (2023) **Dopaminergic signals for reward, performance and social outcomes are dynamically gated during courtship** <https://doi.org/10.1101/822817>
- Sakata JT, Brainard MS (2009) **Social context rapidly modulates the influence of auditory feedback on avian vocal motor control** *J Neurophysiol* **102**:2485–2497
- Sakata JT, Brainard MS (2008) **Online contributions of auditory feedback to neural activity in avian song control circuitry** *J Neurosci* **28**:11378–11390

- Scarpa GB, Starrett JR, Li G-L, Brooks C, Morohashi Y, Yazaki-Sugiyama Y, Remage-Healey L. (2022) **Estrogens rapidly shape synaptic and intrinsic properties to regulate the temporal precision of songbird auditory neurons** *Cereb Cortex* :3401–3420
- Scharff C, Nottebohm F. (1991) **A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning** *J Neurosci* **11**:2896–2913
- Shea SD, Margoliash D. (2010) **Behavioral state-dependent reconfiguration of song-related network activity and cholinergic systems** *J Chem Neuroanat* **39**:132–140
- Singh Alvarado J, Goffinet J, Michael V, Liberti W, Hatfield J, Gardner T, Pearson J, Mooney R. (2021) **Neural dynamics underlying birdsong practice and performance** *Nature* **599**:635–639
- Skals N, Anderson P, Kanneworff M, Löfstedt C, Surlykke A. (2005) **Her odours make him deaf: crossmodal modulation of olfaction and hearing in a male moth** *J Exp Biol* **208**:595–601
- Tumer EC, Brainard MS (2007) **Performance variability enables adaptive plasticity of “crystallized” adult birdsong** *Nature* **450**:1240–1244
- Ulanovsky N, Las L, Nelken I. (2003) **Processing of low-probability sounds by cortical neurons** *Nat Neurosci* **6**:391–398
- Williams AH, Kim TH, Wang F, Vyas S, Ryu SI, Shenoy KV, Schnitzer M, Kolda TG, Ganguli S. (2018) **Unsupervised Discovery of Demixed, Low-Dimensional Neural Dynamics across Multiple Timescales through Tensor Component Analysis** *Neuron* **98**:1099–1115
- Woolley SC, Rajan R, Joshua M, Doupe AJ (2014) **Emergence of context-dependent variability across a basal ganglia network** *Neuron* **82**:208–223
- Xiao L, Chattree G, Ocos FG, Cao M, Wanat MJ, Roberts TF (2018) **A Basal Ganglia Circuit Sufficient to Guide Birdsong Learning** *Neuron* **98**:208–221
- Xiao L, Priest MF, Kozorovitskiy Y. (2018) **Oxytocin functions as a spatiotemporal filter for excitatory synaptic inputs to VTA dopamine neurons** *Elife* **7** <https://doi.org/10.7554/eLife.33892>
- Yip PK, Schmitzberger M, Al-Hasan M, George J, Tripoliti E, Michael-Titus AT, Clayton D, Priestley JV (2020) **Serotonin Expression in the Song Circuitry of Adult Male Zebra Finches** *Neuroscience* **444**:170–182
- Zann RA (1996) **The zebra finch: a synthesis of field and laboratory studies**
- Zhang SX, Miner LE, Boutros CL, Rogulja D, Crickmore MA (2018) **Motivation, Perception, and Chance Converge to Make a Binary Decision** *Neuron* **99**:376–388

Article and author information

Caleb Jones

Department of Neurobiology and Behavior, Cornell University, Ithaca, NY 14853, USA,
 Department of Biomedical Engineering, Cornell University, Ithaca, NY 14853, USA
 ORCID iD: [0000-0001-5894-0239](https://orcid.org/0000-0001-5894-0239)

Jesse H. Goldberg

Department of Neurobiology and Behavior, Cornell University, Ithaca, NY 14853, USA

For correspondence: jessegoldberg@gmail.com

ORCID iD: [0000-0003-0422-8902](https://orcid.org/0000-0003-0422-8902)

Copyright

© 2023, Caleb Jones & Jesse H. Goldberg

This article is distributed under the terms of the [Creative Commons Attribution License \(https://creativecommons.org/licenses/by/4.0/\)](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

Editors

Reviewing Editor

Andrew King

University of Oxford, United Kingdom

Senior Editor

Andrew King

University of Oxford, United Kingdom

Reviewer #1 (Public Review):

Summary:

This study examines the context-dependent modulation of auditory cortical neurons in response to expected sensory input, either self-generated sounds or expected perturbations of self-generated sounds. Specifically, using songbirds, the authors ask whether social context (the presence of a female conspecific) affects 1) the response of auditory cortical neurons to the bird's own song when he is singing; and 2) the response of neurons to perturbations of auditory feedback that the bird has been trained to expect.

Strengths:

First, the authors report that across the population, the responses of the neurons does not differ when a male bird sings alone or if he sings to a female. A fraction of auditory cortical neurons, however, do show significant differences in the firing rate, precision, and/or degree of burst firing when males sing alone vs. when they sing to females. This finding is broadly consistent with the literature showing that sensory neurons (visual, auditory, somatosensory, etc.) can be rapidly reconfigured into different "information processing modes" depending on behavioral state (e.g. quiescence vs vigilance).

For the perturbation experiments, the authors trained birds to expect distorted auditory feedback during a particular syllable. They found that some neurons showed greater responses during perturbation when a female was present (compared to when males were alone) while other neurons had smaller responses during perturbation when a female was present. In addition, the response of a small number of auditory cortical neurons were not affected by behavioral state. These results contrast with their prior report that the responses of midbrain dopaminergic neurons that project to the basal ganglia are "uniformly reduced" in the presence of a female, raising a question of how an evaluation signal is transformed in the circuit from the primary sensory region to the midbrain.

Weaknesses:

While the experiments and analysis are solid, the finding that social context can alter responses of auditory cortical neurons in a multitude of ways (increase, decrease or no change) raises several questions that can be examined with additional analysis. For example, do context-dependent differences in auditory responses derive from context-dependent differences in the songs? Are context-dependent differences present in all classes of neurons and throughout the auditory system?

The observed heterogeneity in the firing properties of auditory cortical neurons, both in response to self-generated sounds and during perturbations of auditory feedback, raises the question of which neurons are sensitive to social context (which likely can be addressed by the authors in a revision). The authors should provide additional details about the recordings:

a) What are the locations of the recording sites?

Prior work has shown that there is an organized map of spectrotemporal features of sounds in the auditory cortex of songbirds; spectral tuning widths change along the medial-lateral axis and temporal tuning widths differ between the input and output layers of Field L. Were the recordings primarily in Field L2 (thalamo-recipient region), L1 or L3? Were some recordings lateral to Field L in secondary auditory regions? Were the neurons that showed context-dependent changes in firing properties localized or distributed throughout Field L (i.e., were the context-dependent differences in neural responses truly brain-wide)? At a minimum, the authors should include a schematic showing the different regions of Field L and a summary of the location of the recording sites. Images of the processed tissue with electrolytic lesions would also be helpful.

b) Was the context-dependent modulation limited to a particular class of neurons (distinguished by spike waveform shape, spontaneous firing rate, or other feature)?

While the authors attribute differences in the responses of single auditory cortical neurons to the presence of a female, other potential explanations for the observed differences should be examined (and potentially ruled out):

a) Prior work has shown that songs of zebra finches differ slightly when males sing alone compared to when they sing to females: songs are faster; pitch is less variable; and the number of introductory elements is greater when males sing to females. Do some of the observed social context-dependent differences in the responses of auditory neurons reflect differences in the songs in the two conditions? This idea is supported in part by a prior study in juvenile zebra finches (Keller & Hahnloser, 2009) showing that ~20% of the neurons they recorded in Field L and a secondary auditory region (CLM) showed anticipatory activity even before the onset of a song bout, suggesting a source of premotor (or at least non-auditory drive) to neurons in the auditory cortex. Did the authors of this study also find premotor activity in Field L, and if so, did it differ between the two social contexts? Might differences in Field L responses reflect motor/song differences?

b) For the perturbation experiments, the authors report heterogeneous responses to playback, with some neurons firing more and other firing less when a female is present compared to when the male is alone. Keller and Hahnloser (2009) found that in juvenile birds, responses of Field L to perturbations of auditory feedback were sensitive to sound amplitude; perturbation responses increased with relative perturbation amplitude. This raises a question of whether perturbation amplitude is different when a male is alone and when a female is present (i.e., the male may move towards the female when she is present and if the speaker is close to the female, the perturbation may be louder than when the male is alone; alternatively, the male be more active when he is alone so the loudness of the perturbation may be more variable across song bouts). It would be useful to know if (and

how much) perturbation amplitude varied depending on the location inside the cage as well as whether the sound pressure level of the underlying song was higher (e.g., Lombard effect). Addition of details of the experimental setup/procedure would help to allay concerns that the amplitude of the white noise varied significantly depending on behavioral context.

Finally, I am still trying to make sense of the differences in the context-dependent modulation of responses of auditory cortical neurons vs. midbrain dopaminergic neurons. Given the heterogeneity of responses in Field L, both to self-generated sounds and to expected perturbations during singing, how are the signals decoded downstream of Field L? At the population level, neither the mean firing rate nor the timing of firing of Field L neurons changed with courtship. Similarly, across the population, the responses to perturbations of auditory feedback were not affected by courtship state (error signal attenuated in 11 neurons, increased in 22 neurons and not affected in 10 neurons). Yet, the courtship state "uniformly" reduces the response of midbrain dopaminergic neurons to auditory perturbation. It would be helpful if the authors could include a model and/or more discussion of how this change may arise.

Reviewer #2 (Public Review):

Summary:

In the manuscript 'Auditory cortical error signals retune during songbird courtship', Jones and Goldberg study auditory cortex in male zebra finches. They explore song-related responses in two different contexts, when the male is either alone or in the presence of a female. Social-context related responses are hypothesized based on previous results on downstream VTA neurons where such modulation is found. They play jamming stimuli through a loudspeaker to probe sensitivity of song-related neural responses to these external stimuli. They find a heterogeneity of responses, in line with auditory cortical neurons computing the social modulation of responses found in VTA.

Strengths:

In general, the work is interesting and sheds light onto auditory processing and self-perception mechanisms in songbirds.

Weaknesses:

Stability of responses has not been studied: some neurons seem to have responses that slowly drift in time, which could lead to observed differences between alone and with-female conditions. Also, possible motor confounds and sound-of-audience confounds should be addressed. The language is often imprecise.

Stability and Reversal: It is a bit unfortunate that stability of effects seemingly has not been studied by reversing experimental conditions. The work would be much stronger if authors could show that audience-dependent tuning is robust in individual cells. Did they record from some neurons during reversal back to the alone condition? Ideally, the responses should be identical before and after recording with an audience. This would control for possible non-stationarities in their neuron recordings/spike-sorting/circadian trends. If authors do not have such data, it would be worth while to even just try to divide the dataset for each neuron and condition (either the audience or isolate condition) into two parts to verify that the response is the same in either part (provided sufficient song renditions are recorded). See also my comment below about Fig. 2A.

Motor responses: Does DAF playback change song? If so, especially if it applies only in one of the two conditions (audience/no audience), then the observed response differences could be motor-related rather than auditory responses. Analyses of song spectrograms right after DAF would presumably provide the answer.

Similarly, motif-aligned spiking activity was time warped to the median duration of undirected or directed motifs. Could the shorter motifs during directed song (as has been reported in other studies) lead to alignment differences that would account for the different error responses in alone/wfemale conditions? In other words, could increased error responses be due to the fixed 100 ms analysis window of the audience condition that extends into a song region beyond the 100 ms region of the no-audience condition where there is increased firing? And vice versa for observed decreases in error responses, i.e. is there a firing pause just after the offset of the 100 ms window in the no-audience condition that causes audience dependence of responses? A simple compensation of song tempo differences by shortening/stretching the analysis window in one of the two conditions would allow to test for this.

Audience versus sound of audience: In the first sentence of the discussion authors write: we discovered that auditory representations of an animal's own vocalizations change with an audience. Is it truly the audience that causes the difference in error responses or is it the sounds the audience makes? To control for that would be to play back stimuli that simulate a non-silent audience through a loudspeaker to see whether error responses depend on the soundscape created by a typical audience (either present or absent). Authors probably do not have such data and to record it would go beyond the scope of this study, but it would be important to discuss this possibility or perform some analysis in that vein.

Reviewer #3 (Public Review):

Summary:

In this study, Jones et al. examine how neural activity in a primary auditory area (field L) of singing male songbirds is modulated by the presence or absence of an audience (a female conspecific). Prior work has demonstrated that the presence of an audience attenuates the responses of dopaminergic neurons to distortions of auditory feedback (DAF). Here the authors report that even in a region that is primarily considered sensory, responses to DAF are also modulated by the audience, although in a heterogeneous manner that does not readily explain previously observed attenuation. These findings address an interesting question and will potentially be important in adding to an understanding of how non-sensory factors can alter response properties of neurons even in primary sensory regions in a context dependent fashion. However, to be fully persuasive, additional analyses will be required to address how much of the apparent modulation by audience may be explained by other factors such as changes in recorded neurons or their properties over time.

Full Public Review:

In this study, Jones et al. examine how neural activity in a primary auditory area (field L) of singing male songbirds is modulated by the presence or absence of an audience (a female conspecific). They test whether activity in Field L differs between conditions in which the male is singing to a female (directed song) or alone (undirected song) and whether response to distortions of auditory feedback (DAF) differ between these conditions. Previous work has shown that in other parts of the songbird brain, sensory-motor activity can differ between directed and undirected song, and that responses to DAF are attenuated when males sing directed song versus undirected song. These prior results raise the interesting question of the extent to which such modulations of activity by the presence of an audience are already present in primary sensory areas such as Field L. This possibility is also motivated by prior work that has shown that Field L activity is not exclusively explained by auditory input, but can also be modulated by the bird's state - whether it is singing or not.

Against this background, the questions asked here are of interest for two inter-related reasons:

1. the authors address whether the presence of an audience (a female conspecific) alters activity in a primary auditory area during singing. Primary auditory areas such as Field L, and analogous mammalian thalamo-recipient cortical regions such as A1, are often thought of as responding very specifically to the features of sensory stimuli, but are also understood to be modulated by a variety of factors including the attentional and behavioral state of the animal. For audition, such modulation includes whether or not animals are vocalizing and listening to themselves or listening to playback of their own vocalizations. Cited works from Keller (2009) as well as Eliades and Wang (2008) have indicated that the act of vocalizing can modulate auditory responses to self-generated feedback in primary auditory areas relative to those arising from playback of the same sounds. Here, the question is whether responses to self-generated feedback differ between conditions of singing alone versus singing to a female audience. A demonstration that the presence of an audience matters to responses in Field L would add to a general understanding of how it is that non-auditory factors can modulate sensory responses.
2. the authors address the possible source of an audience-dependent modulation of responses to feedback perturbation in the VTA previously reported by Goldberg and colleagues (2023). In the VTA, responses to perturbations during singing are consistently attenuated when males are singing to females versus when they are singing alone, but the underlying mechanisms of this modulation are unknown. Here, the authors test the possibility that such modulation by an audience is already present at the level of Field L. The previously reported attenuation in VTA is quite striking and reflects a nice example of how neural processing can differ with varying behavioral priorities. Understanding whether this modulation of responses to DAF arises already in primary auditory areas would further a mechanistic understanding of an intriguing example of state-dependent modulation of sensory processing and behavior, and lend broad insight into related phenomena.

The authors report 1) that activity in Field L differs between directed and undirected singing at many individual recording sites, but that these changes are heterogeneous, with both increases and decreases in activity, so that there is no consistent change across the population and 2) that the responses to DAF differ between directed and undirected song, but that there is no consistent attenuation of response (as observed in the VTA) and instead heterogeneous increases and decreases in response to DAF so that there is no net change at the population level.

These findings, if firmly established, are important and of general interest. While they do not readily explain the source of the audience-dependent attenuation of auditory responses to DAF in the VTA, the demonstration of audience-dependent modulation of self-generated feedback and its disruption in a primary auditory area is an exciting result that would provide an opportunity for further investigation of how changes in social context influence brain and behavior. The manuscript is generally well written, although the presentation is terse. My main reservations about the current manuscript relate to aspects of experimental design and analysis that need to be clarified and addressed before these conclusions will be fully persuasive. There are also some places where further discussion of the findings and their relationship to prior studies would be helpful.

1. A central concern relates to whether the main reported effects associated with differences in singing directed versus undirected song reflect only those changes in conditions, versus contributions from changes in unit isolation or response properties over time. The authors record undirected song in a block in the morning and only after collecting at least 40 renditions do they later record responses during directed song over a series of repeated exposures to a female. Therefore, differences between data collected during undirected song

and directed song also reflect differences between data collected initially during the morning versus later. It is unclear from methods whether any of these recordings during undirected and directed conditions are interleaved, but if this is not the case, then it is crucial to ask how stable were neural recordings with respect to unit isolation, and potential changes to response properties, over the duration of the experiments. This would be less of a concern if the results mirrored those observed in the VTA, where attenuation of responses was observed across the entire population during directed versus undirected conditions - it is hard to explain a phenomenon that is consistently observed across the population as arising from a change in which neurons and spikes are contributing to responses, or other forms of non-stationarity. However, because there are no significant differences reported at the population level in the current study, it is important to address the possibility that observed differences between conditions reflect some form of noise or drift in recorded units, rather than being entirely due to directed versus undirected singing. I have elaborated in more detail below on this concern, including places where the data seems to suggest some non-stationarity of responses, and have some suggestions for ways in which this concern might be addressed.

2. A second concern, related to this first one, has to do with the categorical definition of 'error neurons'. The authors note in their text that it could be problematic to apply categorical definitions to continuous distributions, and yet that seems to be what they then do. The authors have a metric of error sensitivity that they apply to each neuron's response to DAF in both undirected and directed conditions (the error score). They show that there is a continuous distribution of error scores (Figure 2 - figure supplement 1) across the population, with no bimodality that would be suggestive of distinct error sensitive and error-insensitive neurons. One nice feature of their analysis is that they also show the distribution of error scores computed in an analogous fashion for a period of neural activity in the song prior to DAF. This control data set makes it persuasive that there is a significant response to DAF, but also shows that there can be a broad range of error scores even when no DAF has been played, and that this range of 'noise' responses to DAF overlaps substantially with the actual responses to DAF. Despite the continuum of error scores, the authors define a subset of neurons as error responsive only if their responses to DAF exceed a specific threshold (2.5 standard deviations). One of the main conclusions of the paper is based on finding a subset of 22 neurons that exhibited error responses (by this definition) only during singing to a female and 11 neurons that exhibited error responses only when singing alone. These neurons are described as 'retuned' because they have error responses in only one condition.

The problem here is that for some, if not many, of the neurons that are categorically defined as being responsive to DAF in only one condition (directed versus undirected) there is almost certainly not a significant difference in the actual responses to DAF between conditions. This is apparent in the relevant data figure (figure 2 - figure supplement 1) and is a consequence of using a threshold to split a continuous distribution into groups defined as error responsive or not. For example, several neurons in this plot that have almost identical scores in the directed and undirected condition are counted as examples of retuning because the error scores are just a bit over 2.5 in the directed condition and just a bit under 2.5 in the undirected condition.

That this kind of categorical approach may be problematic is apparent in the control data in the plot. Despite the absence of any perturbation, there are error responsive neurons present in these data that are considered selective for directed versus undirected singing - this is an expected consequence of using a threshold on dispersed or noisy biological data. Shifting to a more stringent threshold of three standard deviations, as the authors do, does not help with this problem, as that still treats as categorically different responses that fall on either side of a line, even if only by a tiny amount. I suggest that the authors devise a measure for each neuron to test whether the responses to DAF are significantly different under the two conditions (directed versus undirected). As noted above, this measure should take into account some assessment of the stationarity of responses, as well as the distribution of

responses (which, in some of the examples does not seem to be Gaussian around a mean response level, but rather highly variable across trials).

3. There are several places where further discussion of the previous literature and how the current results relate to that literature would be helpful. This includes:

3a. Some discussion of what is already known about the auditory tuning of field L, and the extent to which responses associated with distortion of feedback may reflect the frequency tuning of field L neurons versus something that might be construed as more specifically as detecting an error in perceived feedback. For example, Field L neurons have previously been characterized as having relatively simple spectro-temporal receptive fields, often with a single frequency band that is excitatory and nearby frequency bands that are inhibitory. It would be beyond the scope of this paper to directly assess the extent to which both song responses and responses to DAF are well predicted by simple STRFs that might be measured for the recorded neurons, or computed from activity during a range of vocalizations, but perhaps worth discussing whether a neuron with such frequency tuning would potentially exhibit 'error responses' of the sort described here, simply because the DAF stimulus happens to fall into the excitatory or inhibitory regions of the neuron's receptive field. While it is OK to use the term 'error responsive' in the current study, it would be good to make clear that changes in firing associated with playing DAF should be expected even for neurons that have simple auditory receptive fields (i.e. with center surround tuning to specific frequencies in a tonotopic map, as has been described for Field L) without necessarily indicating that these neurons are specifically registering any deviation or 'error' between expected feedback and experienced feedback. In this respect, there are multiple subdivisions of Field L with different tuning properties. Please specify further what criteria were used to determine recording locations and how these correspond with previously defined subdivisions.

3b. It would also be useful to discuss further previous work on differences in auditory tuning or responses between conditions when subjects are vocalizing, versus when vocalizations are played back (as in Keller, Eliades) and whether the results in the current study are similar or different. For example, this prior work has indicated that efference copy or other signals that precede vocalizations can reach and influence activity in auditory areas - with the most compelling evidence for this being the modulation of activity prior to the onset of vocalizations. Was this also observed in the current study, and to what extent might this kind of mechanism contribute to the processing of feedback distortions? With respect to this kind of efference signal, or other possibilities, can the authors provide some discussion or speculation about possible mechanisms that might be differentially engaged between conditions of singing directed versus undirected song?

3c. The previous study on DAF responses in VTA indicates enhanced responses to female calls during directed song. To what extent did the current study control for any vocalizations or other sounds produced by females during the directed singing, and could this have contributed to differences in Field L activity between conditions? This question is motivated partly by the highly variable responses in raster plots even within one condition - might some of this reflect motifs during which transient noises are produced from female calling or other movements by the male or female?

More regarding stability of recordings:

The data presented in Figure 1D illustrate some of my concerns about the stationarity of recordings. In the directed condition there are no spikes at all following the first handful of motif renditions. Were the directed and undirected recordings interleaved here? If not, could the recorded neuron simply have been lost, changed in amplitude of recorded spikes so that it was no longer counted, or reduced its responsiveness over the course of the recordings? Because the recordings of undirected and directed singing are described as occurring sequentially, it seems likely that this type of change in recorded signal could contribute to

changes in measured responses over time, independently of effects due to directed versus undirected singing.

A minor issue of this example is that the raw example trace with male alone does not seem to have a corresponding set of points in the raster plot. For panel E, I also cannot find rasters that correspond to the example recordings shown at top.

Figure 2A also shows a neuron that looks like it has non-stationarity; for the alone condition without altered feedback, the main peak has no spikes for the bottom half of the rasters. For the directed condition, much of the difference between control and distorted feedback conditions seems to come from a few trials towards the bottom of the raster plot that show more and earlier firing than most other rasters.

Other more subtle examples are suggested in the figures, such as Figure 1F where responses in the alone condition seem to increase over the course of recordings. A related issue apparent in some of the raster plots is that the firing rate distributions within a given condition sometimes appear to be very non-gaussian, with some motifs during which there is a lot of activity, or apparent bursting, and others in which there is little activity. In addition to the examples above, this includes responses in Fig 1E and Fig 2F. Does anything distinguish these cases or trials? Where differences between conditions are driven by firing differences that are present on only a subset of trials, such as in Fig 2A, there is some deviation from the normal criteria for use of T-tests/Z-scores. Please consider this point and discuss any caveats and/or apply other tests (Monte Carlo? Non-parametric?) as appropriate.

These potential issues of non-stationarity, and non-Gaussian firing rate distributions in each condition, make it complicated to think about what differences in activity reflect changes from undirected to directed conditions versus these other factors.

Approaches to addressing this issue could include more specifically indicating examples in which recordings from the alone condition and directed condition are interleaved and exhibit reversible (between conditions) changes in the pattern of responses (both without DAF in comparing alone versus directed, and with DAF demonstrating differences in DAF influences between conditions). Some good interleaved examples of this sort would be very helpful to illustrate the robustness of differences between conditions. More generally, the methods and or raster plots should include some further explanation of the time periods over which recordings were made in the alone versus directed conditions, and the extent to which they are interleaved or not.

Another approach that could be used if there are not many instances of inter-leaved recordings is to try to document the stationarity or stability of unit isolation and/or responses over time. It would be most helpful when applied to recordings from a given singing condition (i.e. alone or directed) that are interleaved, but even in cases where this is not possible perhaps one could assess the stability of waveforms and unit isolation across time. For example in Figure 2 - Supplementary figure 2, the left-hand and middle examples appear to have quite good unit isolation, and might be the sorts of cases where measures of unit isolation and waveform stability could be used to argue that a gain or loss of spikes due to drift in recordings or changes to SNR and spike detection are not contributing to changes in firing patterns over time (and across conditions).

It potentially would also be informative to present the prevalence of the main effects reported in the study as a function of some measures of unit isolation, SNR, and recording stability. It would be reassuring to see that significant differences between conditions are equally or more prevalent under the conditions of greatest unit isolation and recording stability than in cases with worse SNR or stability.

One other way that the authors might be able to address my main concern would be to look at the stability of firing patterns within conditions, where differences across trials most directly indicate the potential contributions of technical or biological changes in neural activity over time that are not related to the experimental conditions.

To further address some of these issues, it would be helpful to have additional explanations in this paper (rather than by reference to Goldberg and Fee, 2010) of the criteria that were used for counting spikes, and assessing stability of recordings. All I found about this in the Goldberg and Fee, 2010 reference was that "Spikes were sorted off-line using custom Matlab software" Does this require human inspection and judgment? Is there a simple threshold, or waveform measurement used for detecting spikes from single units? Are some sort of signal to noise measures, or ISI violations used to score how well units are isolated?

For the specific examples shown in figures, it would be useful to indicate by small tick marks or otherwise which spikes were counted as single units. For example in figure 2 column B, for the condition with female, did only the 1-3 largest spikes get counted, or also the spikes of medium height?

Page 11: "Many channels on the probes recorded multi-unit activity, which were taken note of but not analyzed in this study."

What were the criteria for this? For several of the examples in the figures there are spikes of varying amplitudes and as mentioned above it would be helpful to clarify how the spikes were sorted into single units in such cases.

Categorical scores:

Page 13: "Neurons with error responses greater than 2.5 in only one condition (undirected versus directed) were considered to have retuned; neurons with error scores greater than 2.5 in both conditions were considered not to have retuned."

This definition results in cases where responses of 2.45 vs 2.55 are described as 'retuned', even if these responses are not significantly different. The figure (Figure 2 - figure supplement 1) indicates that multiple neurons that were scored as retuning had responses that fall very near the threshold in this way.

Page 13, "Our results did not fundamentally change with ... a more stringent threshold of 3..."

The stringency is not issue here, rather the categorical threshold. Retuning would be more persuasively demonstrated if the authors could provide a test of whether or not the responses for individual neurons differ significantly between conditions appropriately taking into account multiple comparisons, stability of recordings, non-Gaussian firing rate distributions across motif renditions, etc. and use this metric to report effects, rather than setting a categorical threshold.