Dopamine D1 receptor activation drives plasticity in the songbird auditory pallium

https://doi.org/10.1523/JNEUROSCI.2823-20.2021

Cite as: J. Neurosci 2021; 10.1523/JNEUROSCI.2823-20.2021
Received: 5 November 2020
Revised: 9 March 2021
Accepted: 21 April 2021

This Early Release article has been peer-reviewed and accepted, but has not been through the composition and copyediting processes. The final version may differ slightly in style or formatting and will contain links to any extended data.

Alerts: Sign up at www.jneurosci.org/alerts to receive customized email alerts when the fully formatted version of this article is published.
Title: Dopamine D1 receptor activation drives plasticity in the songbird auditory pallium

Abbreviated title: D1 receptors drive plasticity in songbird pallium

Authors: Matheus Macedo-Lima¹,²,³, Hannah M. Boyd², Luke Remage-Healey¹,²,*

Affiliations: ¹Neuroscience and Behavior Program, ²Center for Neuroendocrine Studies, University of Massachusetts Amherst, Amherst MA, USA 01003. ³CAPES Foundation, Ministry of Education of Brazil, Brasília, DF, Brazil 70040-020.

*Corresponding author: lremageh@umass.edu

Number of figures: 9
Number of tables: 1
Number of pages: 58
Number of words in abstract: 149
Number of words in introduction: 649
Number of words in discussion: 1411

Competing interests: The authors declare no competing interests.

Acknowledgments: We thank current and former members of the Healey Lab at the University of Massachusetts who helped with this project, especially Amanda Krentzel, Catherine de-Bournonville, Christina Moschetto, Daniel Pollak, Daniel Vahaba, Garrett Scarpa, Jeremy Spool, Katie Schroeder, Maaya Ikeda, Marcela Fernandez-Peters and Rachel Frazier. Finally, we thank Geng-Lin Li, Joseph Bergan, Jeffrey Podos, Karine Fenelon, and two anonymous reviewers for valuable input.

Abstract:

Vocal learning species must form and extensively hone associations between sounds and social contingencies. In songbirds, dopamine signaling guides song motor-production, variability, and motivation, but it is unclear how dopamine regulates fundamental auditory associations for learning new sounds. We hypothesized that dopamine regulates learning in the auditory pallium, in part by interacting with local neuroestradiol signaling. Here, we show that zebra finch auditory neurons frequently coexpress D1 receptor (D1R) protein, neuroestradiol-synthase, GABA, and parvalbumin. Auditory classical conditioning increased neuroplasticity gene induction in D1R-positive neurons. In vitro, D1R pharmacological activation reduced the amplitude of GABAergic and glutamatergic currents and increased the latter’s frequency. In vivo, D1R activation reduced the firing of putative interneurons, increased the firing of putative excitatory neurons, and made both neuronal types unable to adapt to novel stimuli. Together, these findings support the hypothesis that dopamine acting via D1Rs modulates auditory association in the songbird sensory pallium.

Keywords: audition, zebra finch, aromatase, vocal learning, adaptation, memory, learning
Significance statement:

Our key finding is that auditory forebrain D1 receptors modulate auditory plasticity, in support of the hypothesis that dopamine modulates the formation of associations between sounds and outcomes. Recent work in songbirds has identified roles for dopamine in driving reinforcement learning and motor variability in song production. This leaves open whether dopamine shapes the initial events that are critical for learning vocalizations, e.g., auditory learning. Our study begins to address this question in the songbird caudomedial nidopallium (NCM), an analogue of the mammalian secondary auditory cortex. Our findings indicate that dopamine receptors are important modulators of excitatory/inhibitory balance and sound association learning mechanisms in the NCM, a system that could be a fundamental feature of vertebrate ascending auditory pathways.
1.1 Introduction

Studying vocal learning in songbirds continues to provide insight into the suite of mechanisms that support spoken language (Jarvis, 2019). The first step in the spoken language learning process is to make socially-contingent associations about complex sounds. This likely engages cortical brain structures, where sounds and meaning are bound.

Midbrain structures that process reinforcement can guide formation of auditory associations in the cortex through dopamine release. In mammals, dopamine in the primary auditory cortex drives auditory learning and plasticity of simple stimuli, such as pure tones (Bao et al., 2001; Schicknick et al., 2012, 2008). Dopaminergic signaling has also been examined in the songbird brain, but primarily in the context of reinforcement learning for song production and motivation (Gadagkar et al., 2016; Leblois et al., 2010; Matheson and Sakata, 2015; Schmidt and Ding, 2014; Tanaka et al., 2018). Much less is known about how dopamine signaling may regulate auditory processing and association of complex sounds such as song.

The songbird caudomedial nidopallium (NCM; Fig. 1A) is a pallial auditory region considered analogous to the center for speech comprehension in humans, Wernicke’s area (Bolhuis et al., 2010). Neuronal responses in awake, restrained zebra finch NCM show stimulus-specific adaptation to sounds played repeatedly, consistent with active memory formation (Chew et al., 1996; Lu and Vicario, 2017). This adaptation in NCM neurons reflects familiarity with songs, as well as song-consequence associations (Bell et al., 2015; Chew et al., 1996; Lu and Vicario, 2017). The NCM is an important target of a wide variety of neuromodulators, e.g. norepinephrine (Ikeda et al., 2015; Lee et al., 2018), neuroestradiol (Macedo-Lima and Remage-Healey, 2020; Saldanha et al., 2000) and dopamine (Matragrano et al., 2012). Moreover, NCM is rich in NMDA receptors (Saldanha et al., 2004), which are classically regarded as key players in cellular memory formation processes, such as long-term depression and potentiation (Lüscher
and Malenka, 2012). NCM therefore is a highly plastic structure involved in the processing, and perhaps association, of complex sounds. However, the molecular and neuromodulatory mechanisms underlying associations between sound and context in higher-order circuits like NCM have not been elucidated.

Dopaminergic fibers permeate the secondary auditory regions NCM and the caudomedial mesopallium (CMM), but not the thalamo-recipient auditory region, Field L (Reiner et al., 1994; von Eugen et al., 2020). Dopamine receptor mRNA maps onto this architecture, such that D1 receptors (D1Rs) are abundant in NCM and CMM, and D2 receptors are abundant in CMM but not in NCM. Neither receptor is evident in the primary auditory pallium Field L (Kubikova et al., 2010). Neurons in the ventral tegmental area have been reported to project to the NCM and could provide a learning-contingent source of dopamine to this region (Yanagihara et al., 2019).

Indeed, hearing song increases rapid turnover of dopamine in songbird auditory regions, especially in the NCM (Matragrano et al., 2012). Therefore, dopamine signaling is a feature distinguishing secondary from primary auditory pallium in songbirds, providing an anatomical circuit-basis for dopamine-dependent auditory learning. The extent to which dopamine signaling modulates the encoding and processing of stimuli in songbird neurons, however, is unclear.

In this study, we hypothesized that dopamine signaling via D1 receptors (D1Rs) mediates NCM neural circuit plasticity. We obtained anatomical, behavioral, and physiological evidence that support this hypothesis. First, we used immunofluorescence to characterize D1R-expressing neurons regarding aromatase, GABA, and parvalbumin expression. To investigate whether D1Rs were involved in auditory association, we examined the expression of the immediate-early gene EGRI in D1R-expressing neurons during an auditory classical conditioning task. To characterize the synaptic effects of D1R activation we employed in vitro intracellular recordings and recorded
glutamate- and GABAergic currents while delivering a D1R agonist. Finally, to examine the functional effects of D1R activation on auditory physiology and plasticity we delivered the D1R agonist during \textit{in vivo} awake extracellular multielectrode recordings and evaluated whether D1R activation modulated neuronal firing patterns and stimulus-specific adaptation in response to auditory signals.

1.2 Material and methods

1.2.1 Animals

A total of 53 adult (> 90 days old) zebra finches (34 males) were used across all experiments. Birds came from the University of Massachusetts Amherst colony (14:10 hour light-dark cycle) and were not actively breeding (single-sex cages). All procedures were in accordance with the Institutional Animal Care and Use Committee at the University of Massachusetts Amherst.

1.2.2 Immunofluorescence, imaging and quantification

Four males and four females were used for immunofluorescence experiments, to characterize the phenotype of D1 receptor-positive (D1R+) cells. Briefly, animals were taken from the aviary, deeply anesthetized, and perfused with ice-cold phosphate buffered saline followed by room-temperature phosphate-buffered 4% paraformaldehyde. Brains were extracted, postfixed in the same fixative overnight, cryoprotected in 30% sucrose and frozen until processing. With a cryostat, 40 μm parasagittal sections were made and sampled in 4 subseries collected in cryoprotectant solution and stored at \(-20 \, ^\circ\text{C}\) until processed. Of the above subjects, tissue from 2 males and 2 females was processed for D1R, aromatase (ARO) and tyrosine hydroxylase (TH) triple immunofluorescence; tissue from all 4
males and 4 females was processed for D1R, GABA and parvalbumin (PV) triple

Sections containing NCM were selected and transferred from cryoprotectant to phosphate buffer (PB) and washed 3x15 min. They were then incubated in 10% normal goat serum (NGS; Vector Labs, Burlingame, CA) in 0.3% Triton-X (Thermo Fisher Scientific, Waltham, MA) in PB (PBT) for 2 h. Primary antibody solutions (Table 1) were prepared in 10% NGS in 0.3% PBT. To confirm antibody specificity, in a subset of sections the D1R antibody was preincubated for 1 h with blocking peptide (Fig. 1B). Sections were incubated with primary antibodies for 1 h at room temperature, followed by 2 days at 4 °C. Then, sections were washed 3x15 min in 0.1% PBT and transferred to secondary antibody solutions (goat polyclonals; Thermo Fisher; 1:200) prepared in 0.3% PBT. Finally, sections were washed in 3x10 min in 0.1% PBT and kept in the same solution in the fridge until mounted (1 or 2 days later) and coverslipped with ProLong Diamond with DAPI (Thermo Fisher).

Images were taken with a confocal microscope (Nikon A1si). First, NCM was localized and a 4x4 large image was taken at 10x magnification. Then, using only the DAPI channel, the microscope stage was digitally controlled and moved to selected locations on the 10x images, at the ventral and dorsal posterior edges of NCM (Fig. 1C). Then, 15 μm (1 μm step size) z-stack images were taken at 60x magnification, starting from the top-most surface of the section. Two sections per hemisphere per animal were imaged. All laser intensities were maintained uniform across all images within experiments.

D1R antibody penetration noticeably decayed at ~5 μm deep into the tissue, therefore only the top 5 μm of each z-stack was quantified. Cell counts were performed by a blinded experimenter using Fiji (ImageJ, NIH). Briefly, color histograms were set individually for each
image so that background was predominantly dark and only strong signals were counted. Only antibody localization around the nuclei (DAPI) was included. Antibody quantification was done using the z-stack, while DAPI quantification was done using the z-max-projection image. Because DAPI is a universal nuclear stain, only cells with large, ovoidal nuclei (presumably neurons) were counted. The antibody stains employed here further increase this confidence because cytoplasm shape and occasionally dendrites were evident. Still, this method likely underestimates the antibody positivity rate, since glial nuclei are more likely to be overrepresented in the DAPI counts, and aromatase protein has been observed in radial glia in songbirds and other species (Alvarez-Buylla et al., 1987; Forlano et al., 2006).

1.2.3 Classical conditioning and immediate early gene expression

1.2.3.1 Stimuli and Behavioral procedure

Twenty adult zebra finches (10 males) were used in this experiment. Animals were individually isolated for ~17.5 hours in a sound attenuation chamber. The cage set up consisted of three horizontal perches evenly spread out parallel to the cage floor (Fig. 2C). Food and water were located near the middle perch. An LCD monitor and a speaker were positioned adjacent to the cage. Behavior was recorded with a camera positioned on the opposite side of the cage from the monitor.

Animals were exposed to two Experimental conditions (Paired vs Unpaired) and two Reference conditions (Song vs Silence). The aim of the Experimental conditions was to test whether auditory-visual classical conditioning would induce EGR1 expression. Silent videos of conspecifics elicit behavioral attention and engagement and can be employed as social reinforcers for zebra finches (Galoch and Bischof, 2007). Furthermore, neither silent videos nor tones are strong drivers of EGR1 expression in NCM, which confers a dynamic range to evaluate
the effects of stimuli combination (Avey et al., 2005; Sanford et al., 2010). Thus, in the Paired condition (3 males, 3 females), animals were exposed to repetitions of a 2-s pure tone (2 kHz; ~65 dB), the conditioned stimulus, followed by a 2-s delay then a 6-s silent video of a male zebra finch singing (6 different video clips were randomized), the unconditioned stimulus. Stimuli were presented 30 times (30-90 s interstimulus interval) over a 30-min window. In the Unpaired condition (3 males, 3 females), animals were presented with the same amount of stimuli (30 videos and 30 tones), but those were presented separately in pseudorandom order (15-45 s interstimulus interval) over a 30-min window. Videos were scaled so that birds in the video were displayed at approximately real sizes. Conspecific song playback after periods of isolation drive robust expression of EGR1 in NCM when compared to silence conditions (Mello and Clayton, 1994; Mello et al., 1992). Therefore, we exposed birds to Reference conditions to provide expression level references to our Experimental groups. In the Song condition, animals were exposed to three different conspecific bouts of songs (each 18-21 s long; ~65 dB) repeated 10 times in pseudorandom order over 30 min. In the silence condition animals were not presented with any stimulus for 30 min.

### 1.2.3.2 Immunofluorescence

EGR1 protein expression in NCM peaks between 1 and 2 h after induction (Mello and Ribeiro, 1998). Therefore, in all conditions, after the presentation of the last stimulus (or after 30 minutes in the Silence group), chamber lights were turned off (to minimize further stimulation) and animals remained in the chamber for an additional 50 minutes before they were retrieved for perfusion (~10 min until PFA). Total time from beginning of exposure to fixative exposure was ~1.5 h.
After perfusion, brains were processed, sectioned, and stained as described above, but with the following differences. A triple immunofluorescence protocol was performed using antibodies against NeuN, D1R, EGR1 and aromatase (Table 1) prepared in 10% donkey serum (Vector Labs) in 0.3% PBT. Donkey polyclonal secondary antibodies (Thermo Fisher; 1:200) were used. Imaging and cell counts were also performed as described above, except NeuN was used as a background stain.

1.2.3.3 Behavioral scoring and analyses

Video recordings were snipped into 15-s clips around stimulus presentations using custom Python code. Only the Experimental conditions (Paired and Unpaired) were analyzed. Clips were scored twice, once for state and once for event behaviors. State behaviors included beak direction (left, right, pointed at screen, or pointed at camera), sleeping, eating, drinking or continuously moving/flying. Event behaviors were not mutually exclusive with states, and included vocalization, singing, feather ruffling, head tilts, hopping, and gaping.

Videos were scored by an experimenter blinded to the subject ID and trial order. Experimental groups could be inferred by the observer because in the Paired group, the video playback shortly followed the tone presentation. Behaviors were scored using JWatcher (Blumstein and Daniel, 2007).

Behavioral data were extracted and processed using custom Python scripts. Timestamps and continuous behavior durations were aligned to stimulus timestamps and quantified as rates (Hz; event behaviors) or percent of time spent (fraction of stimulus duration; state behaviors). Behaviors outside of the stimuli presentation windows were not analyzed. Beak direction behaviors that enabled subjects to see the screen (beak pointed left, right and towards screen) were summed to comprise a new category termed “screen-time”. Conversely, beak direction
opposite to the screen (back of the head facing video), eating, drinking and sleeping were averaged to comprise a category termed “distractibility”.

1.2.4 *In vitro* whole-cell patch clamp

1.2.4.1 Recordings

Sixteen males were used for slice recordings across two experiments. We focused these experiments on males to further explore mechanisms proposed in a previous behavioral study done in males (Macedo-Lima and Remage-Healey, 2020). We note that we did not observe systematic sex differences in the immunofluorescence and *in vivo* electrophysiology findings, but we do not discard the possibility of sex differences.

After swift decapitation, the top of the skull was resected and the head was immediately immersed in a Petri dish filled with ice-cold carbogen-aerated cutting solution (0-Mg$^{2+}$ cutting solution in mM: 222 glycerol, 25 NaHCO$_3$, 2.5 KCl, 1.25 NaH$_2$PO$_4$, 0.5 CaCl$_2$, 34 glucose, 0.4 ascorbic acid, 2 Na$_2$-pyruvate, 3 myoinositol. Standard cutting solution: same as above except 25 glucose and 3 MgCl$_2$; ~320 mOsm/kg, pH 7.4). In the Petri dish, the cerebellum was resected, and the brain was removed from the skull. Then, brain was removed from the cutting solution and placed on an ice-cold Petri dish lid covered with KimWipe (Kimberly-Clark Professional, Irving, TX). The lateral forebrain of both hemispheres was cut parasagittally to yield flat lateral surfaces, and cerebral hemispheres were bisected. The lateral edges of each hemisphere were then dabbed dry on the KimWipe, glued (cyanoacrylate) to the cutting stage and immediately immersed in the vibratome (VT1000S, Leica Biosystems, Wetzlar, Germany) chamber filled with ice-cold carbogen-aerated cutting solution. Slices were cut at 250 or 300 μm starting from the medial edge which contains NCM. NCM does not have clearly defined lateral boundaries, but song-inducible gene expression experiments indicate strong responses that extend ~1 mm
from the midline (Mello and Clayton, 1994), thus only the first three sections (~750-900 μm) were used for recordings. After cutting, slices were transferred to 37 °C carbogen-aerated 0-Mg2+ or standard recording solution (0-Mg2+; in mM: 111 NaCl, 25 NaHCO3, 2.5 KCl, 1.25 NaH2PO4, 2 CaCl2, 28 glucose, 0.4 ascorbic acid, 2 Na-pyruvate, 3 myo-inositol; standard: idem except 25 glucose and 3 MgCl2; ~320 mOsm/kg, pH 7.4). After a 30-minute recovery at 37 °C and a 30-minute stabilization at room temperature, recordings started. All recordings were performed at room temperature. Slices from both hemispheres were used. Recording location within NCM (e.g. ventral vs dorsal) was not carefully noted, but most recordings were done in the ventral NCM. Dorsal NCM cells were only used if ventral NCM cells were not feasible to patch.

Recording pipettes (borosilicate glass) were pulled with a vertical pipette puller (PC-10, Narishige International, Tokyo, Japan) and had a tip resistance of 4-7 MΩ when submerged in the recording solution and backfilled with either K-gluconate- or CsCl-based solution, for excitatory (EPSC) and inhibitory (IPSC) postsynaptic currents recordings, respectively (K-Gluconate-based: in mM: ~95 K-gluconate, 20 KCl, 0.1 CaCl2, 5 HEPES, 5 EGTA, 3 MgATP, 0.5 NaGTP, 20 creatine-phosphate disodium; osmolarity adjusted to ~295 mOsm/kg with K-gluconate; pH 7.4; CsCl-based: in mM: ~120 CsCl, 8 NaCl, 10 TEA-Cl, 0.2 EGTA, 10 HEPES, 2 MgATP, 0.2 NaGTP; osmolarity adjusted to ~295 mOsm/kg with CsCl; pH 7.4). Internal solutions also contained 0.1% AlexaFluor-488-hydrazide (Thermo Fisher) and 0.1% Neurobiotin (Vector Labs) for “online” and post hoc detection (see below) of the cell, respectively.

Cells were identified with an inverted microscope (Eclipse FN1, Nikon, Tokyo, Japan) with DIC optics. Recordings were made inside a Faraday cage using an EPC-10 amplifier and recorded and compensated (series resistance, slow/fast capacitance) with PatchMaster software (HEKA). Liquid junction potential was automatically subtracted. Traces were digitized at 20
14 KHz. Recordings were made in voltage clamp mode at -70 mV. After whole-cell configuration was achieved, cells were allowed to stabilize for 5 min. Then, baseline drug cocktails (bicuculline or DNQX) were delivered and allowed to act for a minimum of 2 min. Drug-containing solutions were gravity-delivered and flow-matched to a ~2 mL/min peristaltic pump (Cole-Palmer MasterFlex L/S) outtake. Tissue chamber capacity was ~1 mL. Recording quality was constantly monitored between recording blocks and recordings were aborted if series resistance rose above 40 MΩ.

Ten males were used for AMPA/kainate/NMDA EPSCs (sEPSCs) recordings. For sEPSC isolation, bicuculline (20 μM; Santa Cruz) was added to the 0-Mg²⁺ recording solution. For some recordings, after bicuculline was delivered and allowed to take effect, a 1-min baseline recording was made, but for 11 out of 18 cells, a 7-min rundown recording followed. Then, ±-SKF-38393 hydrochloride (10 or 50 μM; Abcam, Cambridge, MA) (Ding and Perkel, 2002) was delivered and a 7-min recording was timed to the start of the delivery. Maximum concentration of drug is estimated to have been achieved within 1 min. Finally, a 10-min washout recording was timed to the start of SKF-38393 clearance, and complete washout is estimated to have been achieved within 1 min. Finally, to confirm the nature of the currents, D-AP-5 (50 μM; Abcam, Santa Cruz, Dallas, TX or Alomone, Jerusalem, Israel) was delivered and currents were monitored for ~5 min.

Six males were used for spontaneous GABA IPSCs (sIPSCs) recordings. For sIPSC isolation, DNQX (20 μM; Tocris) was added to the standard recording solution. Recordings were made similarly to EPSC recordings, except different cells were used for SKF (n=7) and rundown experiments (n=4). After a 1-min baseline recording, either SKF-38393 (10 μM) or nothing (rundown) was added to the recording solution and currents were monitored for 7 min. Then, a
10-min washout (or continued rundown) followed. Finally, to confirm the nature of the currents, bicuculline (20 μM) was added to the recording solution and currents were monitored for ~5 min.

After recordings were completed, the recording pipette was slowly retrieved, and slices were drop-fixed overnight in 4% paraformaldehyde in PB. Then, they were transferred to cryoprotectant solution and kept at -20 °C until processed.

1.2.4.2 Analyses

Recordings were analyzed in IgorPro 6 (WaveMetrics, Lake Oswego, OR). All traces were downsampled (5x) and lowpass filtered at 500 Hz.

In sEPSC experiments, for amplitude measurements (n = 9 from 4 animals; 4 from the left (3 animals) and 5 cells from the right (3 animals) hemispheres), cells were only included in the analysis if series resistance did not change by more than 20% from baseline values. For frequency recordings, all cells [n = 18 from 10 animals; 6 cells from the left (5 animals), 10 cells from the right (8 animals) hemisphere and 2 unnoted cells (2 animals)] were analyzed, as recording quality fluctuations are not expected to interfere with their detection, due to their high amplitude (>50 pA; noise band ~5 pA). Currents were thresholded and manually curated with NeuroMatic (Rothman and Silver, 2018). After curation, currents were automatically measured by custom IgorPro code.

For sIPSC recordings, cells whose series resistance changed more than 20% from baseline were excluded from all analyses. We obtained stable recordings of 11 cells (6 animals), 7 of which (from 5 animals) received the SKF treatment. Those consisted of 2 cells from the left (2 animals) and 5 from the right hemisphere (5 animals). For each cell, one template current was manually selected and spontaneous sIPSCs were automatically detected using a spontaneous
current detection algorithm (Clements and Bekkers, 1997) implemented by Dr. Geng-Lin Li for
IgorPro. After detection, all sIPSCs were measured automatically by custom IgorPro code.
For imaging cells containing neurobiotin, slices were washed 3x15 min in PB and
incubated in 10% NGS in 0.3% PBT for 2 h. Sections were then incubated with rabbit anti-
aromatase diluted 1:2000 in 10% NGS in 0.3% PBT for 1 h at room temperature, followed by 2
days at 4 °C. Then, sections were washed 3x15 min in 0.1% PBT and transferred to 0.3% PBT
containing goat anti-aromatase at 1:200 and Streptavidin-DyLight-488 (Vector) at 1:200. Finally,
sections were washed in 3x10 min in 0.1% PBT and kept in the same solution in the fridge until
mounted (1 or 2 days later) and coverslipped with ProLong Diamond with DAPI (Thermo
Fisher). Cells were imaged using a confocal microscope (Nikon A1si) at 20 and 100x
magnification.

1.2.5 In vivo awake head-fixed electrophysiology
1.2.5.1 Headpost implantation and craniotomy surgery
Five males and four females were retrieved from the single-sex cages and implanted with
headposts. Briefly, under isoflurane anesthesia, custom-made headposts were lowered on the top
of the beak and secured with dental cement. Skull markings over NCM (1.1 lateral, 1.4 anterior
from the bifurcation of the mid-sagittal sinus, 45-degree head angle) were performed, large
craniotomies over NCM were made and meninges were resected. During the recordings
(described below), electrode bundles were always inserted medio-caudally to markings (within
0.5 lateral, 1.2 anterior from mid-sagittal sinus). A small craniotomy was made in an
anterolateral part of the skull for the implantation of a silver ground wire using cyanoacrylate.
Craniotomies were sealed with Kwik-Cast (World Precision Instruments, Sarasota, FL), and
birds were allowed to recover from anesthesia. Recordings were performed within 4 days of
surgery.

**1.2.5.2 Retrodialysis-microdrive (RetroDrive) fabrication**

Custom retrodialysis probe-coupled multielectrode drives (RetroDrives; Fig. 6) were
assembled in house. RetroDrives consisted of a circular printed circuit board (PCB; Sunstone
Circuits, Mulino, OR) soldered to a 36-pin connector (A79026-001, Omnetics, Minneapolis,
MN), and a 5 mm 17G guide tube (stainless steel; Component Supply Company, Sparta, TN). A
strand of 28G enameled copper magnet wire (Remington Industries, Johnsburg, IL) was soldered
to the PCB ground. Three polyimide tubes (2x 198 μm; 1x 100 μm diameter) were inserted
through the guide tube, glued side-by-side on the wall of the guide tube and cut to be flush with
the guide tube. Tetrodes were made by twice-folding and twisting polyimide-coated NiCr tetrode
wire (Sandvik, Stockholm, Sweden) with a tetrode spinner (LabMaker, Berlin, Germany).
Tetrodes were inserted through the larger polyimide tubes (4 in each) and pinned to the PCB. A
single reference wire (50 μm polyimide-coated NiCr wire; Sandvik) was inserted through the
smaller polyimide tube and pinned to the PCB. Wires were glued to the top of the polyimide
tubes. A microdialysis cannula (CMA8011085; Harvard Apparatus, Holliston, MA) was glued
adjacent to the polyimide tubes containing tetrodes, such that a minimum of 3 mm of cannula
protruded from the guide tube. Tetrodes were cut with precision scissors (14058-11, FST, North
Vancouver, Canada) to ~0.5 mm from the tip of the cannula. Reference wire was cut to a similar
length at an acute angle. When 1 mm membrane probes (CMA8011081; Harvard Apparatus)
were inserted, the tips of the probe and wires were offset by ~0.5mm. Importantly, the horizontal
distance between probe wires were ~0.2 mm. Finally, tetrodes were gold-plated to 200-250 kΩ
impedance and all wires and pins were covered with liquid electrical tape (Gardner Bender, New
Berlin, WI) and allowed to dry. RetroDrives were confirmed to successfully operate in NCM using baclofen/muscimol delivery to locally silence neurons within minutes in an earlier study (Macedo-Lima and Remage-Healey, 2020).

1.2.5.3 Recording protocol

On the day of the recording, a microdialysis probe was perfused with artificial cerebrospinal fluid (aCSF; described below) using a microinjection pump (PHD2000, Harvard Apparatus). RetroDrive wires were dipped in 6.25% Dil (Thermo Fisher) in 100% ethanol for visualization of electrode tracks. Then, the animal was comfortably restrained and head-fixed. The Kwik-Cast was removed from the craniotomy over one of the hemispheres. Animal and RetroDrive grounds were connected using alligator clips. The microdialysis probe was inserted through the cannula and the RetroDrive was lowered to NCM (~1.5-2 mm from brain surface; mediocaudally to skull markings). Importantly, tetrodes were positioned medially to the probe such that wires were ~0.5 mm lateral and ~1.2 anterior from the stereotaxic zero (midsagittal sinus).

Recordings were made while animals listened to auditory stimuli and aCSF (PRE) followed by SKF-38393 (SKF) followed by aCSF (POST) were infused during song playback to assess within-subject the effects of SKF on responses to auditory stimuli (described in detail below). Recordings were completed within 4 hours of restraint.

Recordings were made from both hemispheres in different days. When recording in the first hemisphere was completed, the craniotomy was resealed with Kwik-Cast and the animal was returned to the home cage. Within 2 days, the second hemisphere recording was made, after which the animal was overdosed with isoflurane and decapitated. The brain was drop-fixed and
cryoprotected in 30% sucrose in 10% formalin, and frozen until cutting. Cryostat sections were obtained at 40 μm and imaged to confirm location of wires and probe. Recordings were amplified and digitized by a 32-channel amplifier and evaluation board (RHD2000 series; Intan Technologies, Los Angeles, CA) and sampled at 30 kHz using Intan or Open Ephys software. An Arduino Uno (Arduino, Sommerville, MA) was connected to the recording computer to deliver TTL pulses to the evaluation board’s DAC channel bracketing the beginning and end of the audio stimuli (described below) to optimize detection during analysis. Audio playback and TTL pulses were controlled by a custom-made MatLab (MathWorks, Natick, MA) script which also controlled the Arduino and sent a copy of the audio analog signal to the evaluation board ADC channel.

1.2.5.4 Stimuli

Zebra finch songs were obtained from multiple databases (http://ofer.sci.ccny.cuny.edu/song_database), therefore unlikely to have been familiar to our subjects. Twenty-four song files from unrelated birds were bandpass filtered at 0.5-15 kHz and trimmed to include two consecutive motifs without introductory notes in Adobe Audition (Adobe) and mean amplitude-normalized to 70 dB in Praat (Boersma and van Heuven, 2001). Songs were randomly and equally split into two sets, then into 3 subsets containing 4 songs each. For each animal, 1 set was used per hemisphere and, within a hemisphere recording, 1 subset was used per treatment. This was done to ensure that for each treatment birds listened to novel stimuli, because NCM neurons exhibit stimulus-specific adaptation (Chew et al., 1996). Importantly, there was no difference in neuronal firing rates to different subsets, controlled by each stimulus and within each subject and each neuron (GLM/ANOVA: Subset: χ²(5)=5.173,
p=0.395). Therefore, responses across treatments are comparable as they were presumed to reflect responses to novel stimuli.

Each playback session consisted of 4 conspecific songs, repeated 30 times each in pseudorandomized order. Interstimulus interval was pseudorandom within the interval 5±2 s. The speaker was positioned ~30 cm from the animal. Sound level was amplified to ~65 dB as measured by a sound level meter at the animal’s position (RadioShack, Fort Worth, TX).

Playback trial duration lasted ~20 min. Recordings were performed inside a Faraday cage in a quiet room with constant low background noise from a ceiling vent.

Recordings were made from each hemisphere on different days. For each animal’s first recording, the starting hemisphere was randomized in the first subjects, then counterbalanced between sexes. The stimulus set was also initially randomized, then counterbalanced across sexes and hemispheres, but the subset selected for each treatment was always randomized.

1.2.5.5 Drugs and treatment

SKF-38393 (Abcam) aliquots were made in double-distilled water (20 mM; 10 μL) (Schicknick et al., 2012) and kept at −20 °C. On the day of the experiment, one aliquot was added to 1 mL (final concentration 0.2 mM as in (Schicknick et al., 2012)) of previously frozen aCSF aliquots (in mM: 199 NaCl, 26.2 NaHCO3, 2.5 KCl, 1 NaH2PO4, 1.3 MgSO4, 2.5 CaCl2, 11 Glucose, 0.15 bovine serum albumin; pH 7.4). All aliquots were filtered before being loaded into the RetroDrive. Treatment and playback timeline are described in Fig 6. After each round of playback, treatment syringes were switched and a 30 min infusion period elapsed. Retrodialysis speed was set to 2 μL/min (as in Remage-Healey et al., 2010; Remage-Healey and Joshi, 2012; Vahaba et al., 2017).
1.2.5.6 Analyses

Sound playback timestamps were detected using a custom-made audio convolution algorithm in MatLab.

Recordings were highpass filtered at 300 Hz and common-median filtered in MatLab (MathWorks). Single-unit sorting was done with Kilosort (Pachitariu et al., 2016). Sorting results were manually curated in Phy (https://github.com/cortex-lab/phy) and only well-isolated units (high signal-to-noise ratio; low violation of refractory period (the number of interspike intervals within a 1 ms refractory period was 0.25% median in our units); low contamination with other units; segregation in waveform PCA space) were used.

After sorting, for each single-unit, 2000 waveforms were selected pseudorandomly and measured (peak-to-peak duration and ratio; Fig. 4a) in MatLab. All further data processing was done in Python and R.

Spontaneous firing rates were calculated using 500 ms preceding each stimulus playback trial. Within each treatment condition, spontaneous and stimulus firing rates were averaged across stimuli and trials. Peristimulus time histograms (PSTHs) were generated using 10-ms time bins.

Z-scores were calculated by the formula $Z = \frac{\text{Mean}(S) - \text{Mean}(B)}{\sqrt{\text{Var}(S) + \text{Var}(B) - 2 \text{Cov}(S,B)}}$, where $S$ and $B$ are the stimulus and spontaneous firing rates across stimulus trials, respectively. After computing $z$-score by stimulus, those were averaged to yield a single $z$-score per unit per treatment.

Adaptation rates were calculated using trials 6-25, which is the approximate-linear phase of the adaptation profile in NCM (Phan et al., 2006). For each stimulus, the stimulus firing rate across trials was normalized by the firing on trial 6 (set to 100%). Then, a linear regression was
calculated between trials 6 and 25. For each treatment, the minimum (steepest) adaptation slope
across stimuli was used for each unit.

Latency to respond to stimuli were calculated as in Ono et al. (2016). Briefly, for each
stimulus, 5-ms PSTH were generated and convolved with a 5-point box-filter. The latency to
respond to a stimulus was the time after stimulus onset in which the filtered PSTH rose above 3
standard deviations of the average preceding spontaneous firing period (100 ms). If threshold
was not crossed within 400 ms, that stimulus was excluded from analyses.

1.2.6 Experimental Design and Statistical Analysis

Sample sizes and other variables (sex, hemisphere) are reported above for each
experiment’s section.

All statistical analyses and plotting were performed using libraries for R and Python,
respectively. Summary statistics are shown in mean ± standard error of the mean (SEM), unless
specified.

Our general statistical approach was to perform generalized linear modeling (GLM; using
‘glmmTMB’, ‘DHARMa’ and ‘car’ R packages) followed by ANOVA. Data were initially fitted
with gaussian distributions. Normality of residuals was assessed using DHARMa and Q-Q plot
inspection following the GLM fits. If residuals violated normality, data were refit with other
distributions (Poisson or negative binomial) and residuals were reassessed based on new
distributions. If normality was still violated, data were log-transformed when possible (non-
negative, non-zero data) or rankit-transformed (Bliss et al., 1956) and the fitting process above
was repeated. If residuals were distributed according to GLM distributions, the model was
analyzed by ANOVA using Wald chi-square tests. If data still violated residual diagnostics, non-
parametric one-way analyses were performed (e.g. Kruskal-Wallis followed by Dunn’s post-hoc
tests, Friedman test or Wilcoxon signed rank tests).

For immunofluorescence data, quantifications of the two sections belonging to the same
hemisphere and animal were averaged. Response variables were always percentages of the total
number of cells (DAPI) or neurons (NeuN). Data were analyzed by GLM followed by ANOVA,
with Hemisphere and Area (dorsal vs ventral NCM) as interacting fixed factors, controlled for
Area nested within Hemisphere nested within Subject. In the GABA-D1R-PV analyses, tissue
from the right hemisphere of one female was excluded because NCM could not be confidently
localized (off-plane section). In the EGR1-D1R-ARO analyses, data from one male in the Paired
condition could not be quantified because of bad tissue quality after processing. Although
animals of both sexes were used, we did not have power to detect sex differences. Nevertheless,
no qualitative sex differences were observed. For EGR1 induction analyses, GLM followed by
ANOVA was used, with Hemisphere, Area and Condition (Song vs Silence; Paired vs Unpaired)
as interacting fixed factors, controlled for Area nested within Hemisphere nested within Subject.
To test coexpression proportions, chi-square tests were performed on the total sum of cells across
subjects, hemispheres and areas to form a 2x2 contingency table (e.g. ARO+/ARO– vs
D1R+/D1R–). Pearson’s z-scored residuals were analyzed to obtain corresponding one-tailed p-
values.

For behavioral analyses, data were grouped by stimulus type (audio or video) and
averaged across all 30 trials. Then, we ran GLM followed by ANOVA, with Condition (Paired
vs Unpaired) as fixed factor, controlled by Subject.
To evaluate whether behaviors predicted EGR1 expression we fit GLMs including
Hemisphere, Area and Condition (Paired vs Unpaired) as interacting fixed factors and added the
behavioral measurement (e.g. screen-time) as an interacting covariate, controlled for Area nested within Hemisphere nested within Subject. To test the effects of behavior on EGR1 expression with increased statistical power, for every combination of the hemisphere*area interaction we ran individual GLM/ANOVA with Condition and behavioral measurement as interacting factor and covariate.

For patch clamp recordings, minute-by-minute-binned data were analyzed by GLM followed by ANOVA, using Time bins (60 seconds) as a fixed factor controlled over Cell nested within Subject, and Treatment or Dose as independent fixed factor. Bin 1 corresponded to the minute before SKF treatment; bins 2-8 corresponded to the SKF treatment. Washout was excluded from these analyses and is presented for visualization purposes in all figures. Post-hoc Dunnett tests were used to compare treatment bins with the pre-drug bin (control).

For in vivo recordings, cell types were classified using 2-D hierarchical clustering (‘stats’ R package) on peak-to-peak duration vs ratio measurements. The optimal number of clusters was determined using the package Nbclust (Charrad et al., 2014) with the gap-statistic method (Tibshirani et al., 2001). This established technique contrasts the sum of within-cluster distances with distances from clustering a random uniform distribution. The smallest number of clusters that maximizes the distance (gap) between original and random distributions is selected. This method has been used to separate neural waveforms (Nguyen et al., 2015).

After clustering, each unit’s auditory responsiveness was tested by Wilcoxon tests (30x spontaneous vs 30x stimulus trials per song during aCSF infusion). Cells responsive to at least one song were included in all following analyses. Before-drug (PRE) differences among cell types were tested using GLM/ANOVA (controlled over Subject) or Kruskal-Wallis tests and Dunn’s post-hoc tests with Benjamini-Yekutieli false-discovery rate adjustments. Treatment data
were analyzed by GLM followed by ANOVA using Treatment and Cell type as interacting fixed factors, controlled for Unit nested within Subject; Tukey post-hoc tests were used when ANOVA effects with more than 2 factors were significant. To appropriately fit negative-binomial GLMs, firing rate data was summed (instead of averaged) across the 30 stimulus presentations. For a similar reason, to fit gaussian GLMs, z-score data were offset to not contain zero-values ([zscores + 2 × min(zscores)]) and log-transformed. For simplicity, average firing rates and untransformed z-scores are plotted. Due to statistical power limitations, we performed separate analyses excluding Cell type and including Hemisphere and Sex as factors. However, no systematic sex or hemisphere differences with treatment were detected. Washout (POST) was excluded from all analyses.

1.3 Results

1.3.1 Aromatase and D1-receptor proteins are coexpressed by NCM neurons

NCM is a key region for auditory processing in songbirds (Fig. 1A). Previously, we showed that neuroestradiol production in NCM participated in reinforcement-driven auditory association learning (Macedo-Lima and Remage-Healey, 2020). Thus, we hypothesized that reinforcement signaling by the dopaminergic system involved interactions with aromatase neurons in NCM. D1 receptor (D1R) mRNA has been previously reported in NCM (Kubikova et al., 2010), but not its end product, D1R protein. Here we identified a working antibody for the specific detection of D1R protein in songbirds as confirmed by the absence of staining when tissue was preincubated with the D1R antigen (Fig. 1B). We found that D1R+ neurons were often found coexpressing aromatase, representing 29.6 and 35.4% of the aromatase+ (ARO+) neuronal subpopulation in dNCM and vNCM respectively (Fig. 1C). The reciprocal (% of D1R+ neurons expressing ARO) was 26.2 and 32.3% in dNCM and vNCM, respectively. Moreover, the
population of D1R+/ARO+ neurons represented 6.6 and 10% of the neuronal population (DAPI nuclei) in dNCM and vNCM, respectively. A chi-square test analyzing these proportions yielded a significant relationship between D1R and ARO counts ($\chi^2(1)=10.926$, $p<0.001$), with double-labeled cells occurring significantly more frequently than expected (ARO+/D1R+ Pearson’s residual=2.476, $p=0.007$). Interestingly, we frequently observed tyrosine hydroxylase+ (TH+) fibers enveloping D1R+/ARO+ neurons (see Fig. 1D for representative images). These data show that D1R protein is prevalent in NCM neurons and suggest that D1R+ neurons represent a significant population in NCM. Moreover, D1R+ neurons are significantly more likely to express ARO+ than D1R– neurons, suggesting an association between estrogen production and dopaminergic signaling.

1.3.2 Song induction of EGR1 in D1R+ and ARO+ neurons

Next, we examined whether D1R+ and ARO+ NCM neurons were involved in processing of conspecific songs, using the neuroplasticity marker EGR1. Conspecific song playback to birds significantly increased overall EGR1 expression in NCM when compared to birds that experienced silence, which is well established in the literature (Mello and Clayton, 1994; Mello et al., 1992). GLM/ANOVA showed that the degree of song induction of neuronal EGR1 (% of NeuN) depended on the area and hemisphere (Hemisphere*Area*Condition: $\chi^2(1)=7.748$, $p=0.005$). Song exposure increased EGR1 expression in L-vNCM (Tukey’s post-hoc test: Song – Silence: 21.6±4.1 – 8.5±3.5% mean±SEM, $t(18)=2.887$, $p=0.010$) and R-dNCM (Tukey’s post-hoc test: Song – Silence: 53.0±6.3 – 28.1±6.7% mean±SEM, $t(18)=5.012$, $p<0.001$). There was also a trend for an increase in L-dNCM, but not R-vNCM (Tukey’s post-hoc test: Song – Silence: L-dNCM: 44.8±11.4 – 28.5±2.0% mean±SEM, $t(18)=1.820$, $p=0.085$, R-vNCM: 18.0±3.3 – 12.2±11.4% mean±SEM, $t(18)=1.180$, $p=0.253$).
We next asked whether EGR1 song induction also occurred in D1R+ and ARO+ neurons. We tested this in three separate GLMs: one for each marker and one for the colabeling. In D1R+ neurons, song exposure increased EGR1 expression when compared to Silence (Fig. 2B; Area (dNCM>vNCM): $\chi^2(1)=10.202$, $p=0.001$; Condition: $\chi^2(1)=6.774$, $p=0.009$; $p>0.08$ in all other main effects). The same was true for ARO+ and double-labeled D1R+/ARO+ neurons (Fig. 2B; ARO+: Area (dNCM>vNCM): $\chi^2(1)=56.723$, $p<0.001$; Condition: $\chi^2(1)=13.824$, $p<0.001$; $p>0.061$ in all other main effects; D1R+/ARO+: Area (dNCM>vNCM):$\chi^2(1)=3.978$, $p=0.046$; Condition: $\chi^2(1)=4.491$, $p=0.034$; Area*Condition: $\chi^2(1)=3.613$, $p=0.057$; all other $p>0.190$).

These data show that song stimulation increased EGR1 expression in D1R+, ARO+ and double-labeled D1R+/ARO+ neurons. These results are consistent with the notion that dopamine and neuroestrogen signaling are each involved in auditory responses to song and may interact in that capacity.

### 1.3.3 Classical conditioning increases EGR1 expression specifically in D1R+ neurons

We were interested in whether our novel classical conditioning paradigm would induce expression of EGR1 in NCM neurons, so we compared expression levels between the Paired group (sound followed by video) and the Unpaired group (sound and video presented at random; Fig. 2C-D). We found that overall EGR1 neuronal induction (i.e., total EGR1 as a % of NeuN) was not affected by Condition (not shown; Area (dNCM>vNCM): $\chi^2(1)=24.726$, $p<0.001$; Condition: $\chi^2(1)=2.154$, $p=0.142$; $p>0.187$ in all other main effects). However, the frequency of double-labeled D1R+/EGR1+ neurons was significantly increased by Condition (Fig. 2E; Condition: $\chi^2(1)=6.370$, $p=0.012$; $p>0.150$ in all other main effects). This effect was not observed with ARO+/EGR1+ neurons (Area (dNCM>vNCM): $\chi^2(1)=22.430$, $p<0.001$; Condition: $\chi^2(1)=1.027$, $p=0.311$; $p>0.290$ in all other main effects) or with
D1R+/ARO+/EGR1+ neurons (Condition: $\chi^2(1)=1.593$, $p=0.207$; $p>0.232$ in all other main effects). We noted that D1R+/EGR1+ expression levels for both unpaired and paired groups seemed qualitatively lower than the song playback reference control. However, a GLM including all Reference and Experimental groups only found differences between Unpaired and Song (Fig. 2E; Condition: $\chi^2(1)=20.957$, $p<0.001$; Area: $\chi^2(1)=5.777$, $p=0.016$; $p>0.173$ for all other main effects; Tukey’s post hoc test for Condition: Unpaired – Song: $t(54)=-4.484$, $p<0.001$; Silence – Song: $t(54)=-2.790$, $p<0.0035$; Paired – Unpaired: $t(54)=2.481$, $p=0.074$; all other $p>0.174$). This suggests that classical conditioning increased EGR1 expression in D1R+ to statistically similar levels to song exposure, while silence and unpaired treatments remained low.

We note that we observed a significant interaction between hemisphere and condition on our NeuN counts (Area (dNCM>vNCM): $\chi^2(1)=9.910$, $p=0.002$; Hemisphere*Condition: $\chi^2(1)=6.376$, $p=0.012$; $p>0.068$ in all other main effects. Tukey’s post-hoc test Paired vs Unpaired: L hemisphere: $t(33)=-2.217$, $p=0.034$, R hemisphere: $t(33)=0.226$, $p=0.822$). We ran a GLM/ANOVA including NeuN counts as a covariate to examine its impact on D1R+/EGR1+ frequency. We found that NeuN numbers did not account for the differences we observed above, i.e. that classical conditioning increased the frequency of D1R+/EGR1+ expression in NCM neurons (NeuN: $\chi^2(1)=0.998$, $p=0.318$; Condition: $\chi^2(1)=6.748$, $p=0.009$. $p>0.229$ in all other main effects). We are unable to explain this interaction between Condition and Hemisphere in NeuN expression in our sample. Still, this result has little bearing on our main conclusion that classical conditioning increases EGR1 expression in D1R+ neurons.

Next, we wanted to explore whether the increase in EGR1 expression in D1R+ neurons in our Paired group could be explained by any behavioral parameter. Specifically, we were
concerned that differences in “reward” exposure (i.e. more time looking at the screen) could 
underlie the difference we observed between treatment groups. Indeed, we individually tested all 
behaviors we scored and found that average screen-time (beak directions towards left, right or 
screen; see Methods) was significantly higher in the Paired group (Fig. 2F; GLM/ANOVA: 
$\chi^2(1)=4.820, p=0.028$). Similarly, state behaviors that reflected distractibility towards the video 
(eyes closed, eating, drinking, continuously flying, beak direction towards camera) were higher 
in the Unpaired group (GLM/ANOVA: $\chi^2(1)=10.08, p=0.001$).

To formally test whether screen-time predicted EGR1 expression in D1R+ neurons, we 
performed a GLM including screen-time as a covariate. Indeed, screen-time was a predictor of 
D1R+/EGR1+ expression, but there was a significant interaction between all factors (Fig. 2F; 
Screen-time: $\chi^2(1)=9.816, p=0.003$; Hemisphere*Area*Condition*Screen-time: $\chi^2(1)=4.137$, 
p=0.042; all other p>0.068). Distractibility did not predict D1R+/EGR1+ (Distractibility: 
$\chi^2(1)=2.441, p=0.118$; all other p>0.153).

Because we observed a 4-way interaction with hemisphere*area*condition*screen-time, 
we retested these data using GLMs for each hemisphere*area combination. We found that 
Condition (Paired vs Unpaired) and not screen-time was a significant predictor of D1R+/EGR1+ 
neurons in the L-vNCM (Condition: $\chi^2(1)=4.765, p=0.029$; Screen-time: $\chi^2(1)=2.280, p=0.131$; 
Condition*Screen-time: $\chi^2(1)=0.903, p=0.342$), indicating that associating sounds with 
reinforcement drove EGR1 in D1R-expressing neurons. Conversely, in the L-dNCM, screen-
time (but not Condition) was a significant predictor of EGR1 in D1R+ neurons (Condition: 
$\chi^2(1)=0.008, p=0.929$; Screen-time: $\chi^2(1)=3.926, p=0.047$; Condition*Screen-time: $\chi^2(1)=1.884$, 
p=0.170). In the right hemisphere neither factor nor interactions were statistically significant (all 
p>0.231). ANOVA results are illustrated in Fig. 2G.
We also asked whether EGR1 expression was being driven by classical conditioning in only a subset of D1R+ with respect to ARO presence. For D1R+/ARO− neurons we found an effect of classical conditioning (Condition: $\chi^2(1) = 6.8724, p=0.009$; paired > unpaired).

However, when controlling by screen-time, we found that this effect was driven by screen-time and not by Condition (Hemisphere*Screen-time: $\chi^2(1) = 6.542, p=0.011$; Hemisphere*Area: $\chi^2(1) = 4.867, p=0.027$; all other $p>0.11$). In D1R+/ARO+ cells EGR1 expression was not affected by classical conditioning as stated above, but screen-time did affect its distribution (Screen-time: $\chi^2(1) = 7.492, p=0.006$; Area*Hemisphere*Condition*Screen-time: $\chi^2(1) = 4.447, p=0.035$; all other $p>0.190$). However, 2-factor GLMs controlling by Hemisphere and Area only yielded a trend for Screen-time in the R-dNCM (p=0.056) and R-vNCM (p=0.097) and no other effects (all other $p>0.139$). Therefore, the effects of classical conditioning we observed can be explained by the increase of EGR1 expression in D1R+ neurons that can be either ARO+ or ARO− (both types contributing to the effect).

In summary, screen-time was a strong predictor of EGR1 expression in D1R+ neurons (both ARO+ and ARO−), but this effect was particular to the L-dNCM. Importantly, classical conditioning increased EGR1 expression in D1R+ neurons in the L-vNCM which suggests that association during classical conditioning is hemisphere- and subregion-specific. This nuanced regional modulation might have emerged in part due to the nature of our auditory stimulus (2 kHz tone) and the tonotopic organization of NCM (Müller and Leppelsack, 1985).

### 1.3.4 D1R+ neurons are predominantly GABA+

Next, we wanted to characterize the neurotransmitter profile of D1R+ cells in NCM. We found that the majority (58.7 and 64.2%) of D1R+ neurons are GABA+, and these colabeled neurons represent 21.9 and 26.4% of the neuronal population in dNCM and vNCM, respectively.
The reciprocal was also true, such that 54.6% (dNCM) and 56.5% (vNCM) of GABA+ neurons contained D1R (Fig. 3A). A chi-square test analyzing these proportions showed that D1R+/GABA+ co-expression was significantly more frequent than expected ($\chi^2(1)=319.660$, $p<0.001$; Pearson’s residual=9.775, $p<0.001$). Similarly, we observed that the majority (~55%) of parvalbumin+ (PV+) neurons also express D1R, combining vNCM and dNCM (Fig. 3B), and that this neurochemical phenotype represented ~4% of the entire NCM neuronal population (Fig. 3A). The D1R+/PV+ cell population proportion is also more frequent than expected ($\chi^2(1)=22.288$, $p<0.001$; Pearson’s residual=3.470, $p<0.001$). These results show that D1R+ neurons are predominantly GABA+ and represent a significant subpopulation in NCM. Of note, our data show that the majority of PV+ neurons express D1Rs, which suggests this subpopulation is of particular interest for dopamine modulation of auditory processing.

1.3.5 D1R activation reduces the amplitude of GABAergic currents in NCM in vitro

The anatomical findings above led to the hypothesis that activating D1Rs would modulate GABAergic neurotransmission. Thus, we recorded spontaneous postsynaptic currents from neurons in NCM in vitro (Fig. 4A). Inhibitory currents were isolated with the AMPA receptor antagonist DNQX (sIPSCs) and were confirmed to be GABAergic by GABA receptor antagonist bicuculline application at the end of recordings. In separate sets of cells, we either applied 10 μM SKF or nothing (rundown) to the bath. Only recordings with low and stable series resistance throughout the baseline and drug/rundown period (Rs<40 MΩ; less than 20% change; see methods) were included (Rs=16.971±0.826 MΩ, mean±SEM of all recording periods). We obtained stable recordings of 11 cells (6 animals), 7 of which (from 5 animals) received the SKF treatment.
For amplitude measurements, GLM/ANOVA analyses comparing treatments (SKF or rundown) detected a significant interaction between Time and Treatment (Fig. 4B; Time: $\chi^2(7) = 27.781, p < 0.001$; Treatment: $\chi^2(1) = 0.085, p = 0.770$; Time*Treatment: $\chi^2(7) = 15.592, p = 0.029$). Dunnett’s post-hoc tests revealed a significant reduction in sIPSC amplitude only in SKF treated neurons (vs before-treatment: $p < 0.05$ between minutes 4-7 of SKF). For frequency measurements, GLM/ANOVA analyses did not detect significant differences between SKF-38393 and rundown, but detected a significant reduction of sIPSC frequency over time (Fig. 4B; Time: $\chi^2(7) = 27.454, p < 0.001$; Treatment: $\chi^2(1) = 0.096, p = 0.757$; Time*Treatment: $\chi^2(7) = 1.735, p = 0.973$; Dunnett’s post-hoc test vs before-treatment: $p < 0.05$ between minutes 5-7). These results show that SKF-38393 treatment significantly reduced the amplitude of GABAergic currents. The reduction in frequency observed during SKF-38393 treatments did not differ from rundown, indicating that, regardless of treatment, the number of detected sIPSCs decayed over time.

We were unable to test a lateralization effect for this dataset, because after exclusion of unstable recordings, our sample consisted of 2 cells from the left and 5 from the right hemisphere.

**1.3.6 D1R activation reduces the amplitude but increases the frequency of glutamatergic currents in NCM in vitro**

Our anatomical findings suggest that ~38% of D1R-positive neurons are non-GABAergic (i.e. putatively glutamatergic; see Fig. 3A), which led to the hypothesis that activating D1Rs would also directly modulate glutamatergic currents in NCM. Therefore, we recorded excitatory NMDA/AMPA/Kainate currents (sEPSCs) isolated in 0-Mg$^{2+}$ bath containing the GABA$\alpha$-receptor antagonist bicuculline. For amplitude measurements, only recordings with low and stable series resistance throughout the baseline and drug periods ($Rs < 40 \text{ M\ Ω}$; less than 20%
change; see methods) were included (Rs=28.770±2.216 M\(\Omega\), mean±SEM of all recording periods). For frequency measurements, series resistance change was not employed as quality control because the observed currents were very high amplitude and highly unlikely to go undetected with small changes in Rs. Still, only recordings with Rs<40 M\(\Omega\) were included (Rs=26.880±1.428 M\(\Omega\), mean±SEM of all recording periods). We obtained stable recordings from 9 cells (4 animals) for amplitude and 16 cells (8 animals) for frequency.

For these experiments we used two doses (10 and 50 \(\mu\)M) of SKF-38393 in different sets of cells. In these experiments we performed rundown experiments before SKF-38393 treatment in a subset of cells (n=4 for amplitude; n=10 for frequency; see methods). Both amplitude and frequency of sEPSCs were stable during 7 minutes before treatment (Fig. 4C; GLM/ANOVA: Amplitude: \(\chi^2(6)=4.620, p=0.593\); Frequency: \(\chi^2(6)=6.916, p=0.329\)). Currents were confirmed to be NMDA dependent at the end of the recordings.

For amplitude measurements following SKF-38393 treatment, we performed GLM/ANOVA on the effect of different doses of SKF-38393 over time. These analyses showed a reduction in the amplitude of sEPSCs over time due to treatment, and no difference between the two SKF doses (Fig. 4C; Time: \(\chi^2(7)=48.729, p<0.001\); Treatment: \(\chi^2(1)=0.0046, p=0.9461\); Time*Treatment: \(\chi^2(7)=2.950, p=0.890\); Dunnett’s post-hoc test vs before-drug: p<0.05 on minutes 6-7 of SKF). For frequency measurements, GLM/ANOVA analyses showed an increase in sEPSC frequency over time, and again no difference between the two SKF doses (Fig. 4C; Time: \(\chi^2(7)=23.394, p=0.001\); Treatment: \(\chi^2(1)=3.140, p=0.076\); Time*Treatment: \(\chi^2(7)=4.358, p=0.738\); Dunnett’s post-hoc test vs before-drug: p<0.05 on minutes 2, 5 and 6 of SKF).

We also tested whether we could detect a lateralization effect. We recorded from 6 cells from the left (5 animals) and 10 cells from the right (8 animals) hemisphere. For amplitude
measurements, stable recordings were obtained from 4 cells from the left (3 animals) and 5 cells from the right (3 animals) hemispheres (see Methods). Hemisphere was included as an interacting factor with Time (and as a random variable nested within Subject). We did not find significant effects of Hemisphere for either amplitude (Time: $\chi^2(7)=54.660, p<0.001$; Hemisphere: $\chi^2(1)=0.423, p=0.516$; Time*Hemisphere: $\chi^2(7)=10.977, p=0.140$) or frequency (Time: $\chi^2(7)=21.281, p=0.003$; Hemisphere: $\chi^2(1)=2.262, p=0.133$; Time*Hemisphere: $\chi^2(7)=11.943, p=0.102$).

In summary, D1R activation in vitro reduced the amplitude of both GABA and glutamatergic spontaneous currents, while also increasing the frequency of the latter. These findings establish a role for dopamine modulation of network excitation and inhibition in NCM, and predict that D1R activation in vivo causes differential effects depending on cell type (i.e., downregulate vs. upregulate GABAergic and glutamatergic neuron firing, respectively).

### 1.3.7 Cell type separation based on waveform measurements for in vivo recordings

We isolated 107 single-units from 9 adult birds in awake head-fixed recordings using acute recording 32-channel microdrives coupled to retrodialysis probes (RetroDrives; Fig. 6A-B; see methods). We measured peak-to-peak duration and ratio of each unit and analyzed the data using an unsupervised hierarchical clustering algorithm (see methods; Fig. 5A). The gap-statistic results show that the variance in clustering was best explained by 4 clusters, which we named narrow-spiking 1 and 2, and broad-spiking 1 and 2 (NS1, NS2, BS1, BS2 respectively). Cell types differed significantly in waveform peak-to-peak duration (GLM/ANOVA: $\chi^2(3)=299.970, p<0.001$; Tukey’s post-hoc test: all $p<0.01$) and ratio (GLM/ANOVA: $\chi^2(3)=294.880, p<0.001$; Tukey’s post-hoc test: all $p<0.001$ except BS1–BS2 where $p=0.972$). In summary, waveform
peak-to-peak duration followed the pattern NS1<NS2<BS1<BS2 and waveform peak-to-peak ratio followed BS2<BS1=NS1<NS2.

The classification commonly used in the literature of narrow- and broad-spiking neurons in songbird pallium uses only peak-to-peak duration and a division boundary of ~0.4 ms (Aurore et al., 2019; Schneider and Woolley, 2013; Vahaba et al., 2017; Yanagihara and Yazaki-Sugiyama, 2016; but see Bottjer et al., 2019). Using both peak-to-peak duration and ratio, we provide evidence of 2 further subdivisions; therefore, we named our clusters to extend the previous classification: NS1 and NS2 – narrow-spiking; and BS1 and BS2 – broad-spiking.

Following clustering, 15 non-auditory-responsive cells were excluded from the analyses (see methods) and 92 units were further analyzed. This sample consisted of 21 (L) and 23 (R) units from females and 34 (L) and 14 (R) units from males. In this sample, BS2 were more numerous (34.8%) followed by BS1 (27.2%), NS1 (26.1%) and NS2 (12%; Fig 5A).

Representative PSTHs and adaptation slopes of a NS1 and a BS1 are shown in Fig. 5B. The baseline physiological profiles assessed during aCSF infusion showed clear segregation by cell type classification (Fig. 5C). GLM/ANOVA analyses showed that cell types differed in spontaneous firing rates ($\chi^2(3)=50.554, p<0.001$; Tukey’s post-hoc test: p<0.05 in NS1–BS1, NS1–BS2, NS2–BS2 and BS1–BS2), stimulus firing rates ($\chi^2(3)=86.754, p<0.001$; Tukey’s post-hoc test: p<0.05 in NS1–NS2, NS1–BS1, NS1–BS2 and NS2–BS2), z-scores ($\chi^2(3)=36.169, p<0.001$; Tukey’s post-hoc test: p<0.05 in NS1–NS2, NS1–BS1 and NS1–BS2), adaptation slopes ($\chi^2(3)=10.465, p=0.015$; Tukey’s post-hoc test: p<0.05 in NS2–BS2), latencies to respond ($\chi^2(3)=12.166, p=0.006$; Tukey’s post-hoc test: p<0.05 in NS1–BS2) and in the % of songs responded to ($\chi^2(3)=22.785, p<0.001$; Tukey’s post-hoc test: p<0.05 in NS1–BS1 and...
Therefore, the 4 cell types clustered by waveform shape in our recordings also differed in physiological profile. Broadly speaking, NS1 cells had highly symmetrical and narrow action potentials, high firing rates and z-scores, as well as fast response latencies and lower selectivity, which all parallel features of mammalian cortical high-firing PV+ inhibitory interneurons. Compared to NS1, NS2 cells had less narrow and symmetrical waveforms with lower firing rates, which resemble properties of mammalian cortical low-firing somatostatin+ or VIP+ inhibitory interneurons (Tremblay et al., 2016). BS1 and BS2 had broader and less symmetrical waveforms and lower firing rates than NS1/NS2. Additionally, they showed longer latencies to respond to stimuli and higher selectivity, all features that resemble those of mammalian cortical excitatory projection neurons (Atencio and Schreiner, 2008; Hromádka et al., 2008; Tsunada et al., 2012; Wu et al., 2008).

1.3.8 D1R activation reduces spontaneous and stimulus firing of NS1, but increases spontaneous firing of BS1 neurons in vivo

Our *in vitro* experiments generated clear predictions about differential D1R activation on narrow- vs. broad-spiking neurons *in vivo* (see above). We therefore analyzed how SKF-38393 (SKF; 0.2 mM) affected single-unit responses to sound playback (timeline in Fig. 6).

Representative PSTHs for NS1 and BS1 cells are shown in Fig. 7A. We observed effects of D1R activation that were selective for cell waveform phenotypes in NCM.

For spontaneous firing rate, GLM/ANOVA analyses comparing Treatment and Cell-Type showed that SKF reduced the firing of NS1 and tended to increase the firing of BS1 (Fig. 7B; Treatment: $\chi^2(1) = 0.011$, $p = 0.916$; Cell-Type: $\chi^2(3) = 54.694$, $p < 0.001$; Treatment*Cell-Type: $\chi^2(3) = 12.866$, $p = 0.005$; Tukey’s post-hoc test: PRE–SKF: NS1: $t(173) = 2.702$, $p = 0.007$; NS2: $t(173) = 0.712$, $p = 0.477$; BS1: $t(173) = -1.758$, $p = 0.079$; BS2: $t(173) = -1.155$, $p = 0.248$). Because
of the trend we observed, to assess whether a treatment effect could be detected in BS1 units with increased statistical power, we performed a single-factor GLM/ANOVA only on BS1 units, which yielded a significant increase in spontaneous firing due to SKF ($\chi^2(1)=6.008, p=0.014$). Single-factor GLMs for NS2 and BS2 yield non-significant treatment effects (NS2: $\chi^2(1)=0.001$, p=0.972; BS2: $\chi^2(1)=2.650$, p=0.104).

For stimulus firing rate, SKF decreased firing of NS1 cells (Fig. 7C; Treatment:

\[ \chi^2(1)=0.734, p=0.392; \text{Cell-Type: } \chi^2(3)=53.731, p<0.001; \text{Treatment*Cell-Type: } \chi^2(3)=9.043, p=0.029; \] Tukey’s post-hoc test: PRE–SKF: NS1: $t(173)=2.753$, p=0.007; NS2: $t(173)=0.783$, p=0.435; BS1: $t(173)=-1.232$, p=0.220; BS2: $t(173)=-0.183$, p=0.855).

To explore further the change in spontaneous and stimulus firing due to SKF, we plotted a correlation between the %-change in spontaneous versus stimulus firing induced by SKF (Fig. 7D). Values above 0 in either axis indicate an increase in firing due to SKF. Note that on average BS1 and BS2 data points were situated above 0 in both axes, whereas NS1 and NS2 were below 0. Changes in spontaneous vs stimulus firing were also highly correlated (Pearson’s $r=0.876$, p<0.001). Interestingly, the regression line slope’s 95% confidence interval [0.693; 0.875] falls lower than and does not include the slope of the identity line (slope=1), which suggests that spontaneous firing was more affected by SKF than stimulus firing.

For stimulus response z-scores, GLM/ANOVA analyses showed that SKF decreased overall z-scores regardless of cell type (Fig. 7E; Treatment: $\chi^2(3)=8.657$, p=0.003; Cell-Type:

\[ \chi^2(3)=46.855, p<0.001; \text{Treatment*Cell-Type: } \chi^2(3)=4.443, p=0.217). \]

Finally, SKF-38393 treatment increased overall latency to respond (Fig. 7F; Treatment:

\[ \chi^2(3)=5.754, p=0.016; \text{Cell-Type: } \chi^2(3)=13.597, p=0.004; \text{Treatment*Cell-Type: } \chi^2(3)=2.248, p=0.523 \] and decreased the % of songs units responded to (Fig. 7G; Treatment: $\chi^2(3)=8.929$, p<0.001).
Regardless of cell type. Altogether, these results show that the D1R agonist SKF-38393 affects the response properties of NCM cell types differentially. Notably, as predicted by our in vitro experiments, D1R activation reduces the firing of putative inhibitory neurons (NS1) while increasing the firing of putative excitatory neurons (BS1). Furthermore, in all cell types D1R activation decreased z-scores, increased latency to respond and decreased the % of songs units responded to, which might represent direct or indirect consequences of the shift in inhibitory/excitatory balance induced by tonic D1R activation.

1.3.9 D1R activation sharply reduces neuronal stimulus-specific adaptation in vivo

NCM neurons show stimulus-specific adaptation when birds are presented repetitions of the same stimuli, which are thought to reflect short/medium-term memory formation (Chew et al., 1996; Lu and Vicario, 2017). Therefore, we asked whether D1R activation would result in changes in adaptation to novel stimuli. Trial-by-trial stimulus firing rates were used for deriving adaptation slopes (see methods). Eight cells (3 BS1, 5 BS2) had to be removed from the analyses, because firing rate on the first trial used for the regression (trial 6; see methods) was 0 during either PRE or SKF playbacks. Representative rasters and corresponding slopes of an NS1 cell is shown in Fig. 8A. Fig. 8B depicts the slope through the normalized firing rate of trials 6-25 averaged by cell type, for all cell types. GLM/ANOVA analyses showed that SKF-38393 infusion reduced adaptation (i.e. shallower regression slopes), regardless of cell type (Fig. 8C; Treatment: $\chi^2(3)=8.278$, p=0.004; Cell-Type: $\chi^2(3)=20.069$, p<0.001; Treatment*Cell-Type: $\chi^2(3)=5.244$, p=0.155). Note, however that this effect seems to be driven by all cell types but for BS2, which on average appears to remain unchanged by SKF. In fact, BS2 cells were the only
group that retained non-zero slopes during SKF treatment (One-sample Wilcoxon signed-rank tests versus 0; NS1: p=0.308; NS2: p=0.549; BS1: p=0.052; BS2: p<0.001).

Importantly, the % change in slopes did not correlate with the % change in spontaneous firing (GLM/ANOVA; Spontaneous % change: \( \chi^2(1)=1.126, p=0.289 \); Cell-Type: \( \chi^2(3)=2.723, p=0.436 \); Spontaneous % change*Cell-Type: \( \chi^2(3)=3.451, p=0.327 \) or stimulus firing (GLM/ANOVA; Stimulus % change: \( \chi^2(1)=0.800, p=0.371 \); Cell-Type: \( \chi^2(3)=2.764, p=0.430 \); Stimulus % change*Cell-Type: \( \chi^2(3)=2.765, p=0.429 \), suggesting that lower slopes are not predicted by firing rate “floor-effect” (data not shown).

Therefore, D1R activation flattens the stimulus-specific adaptation slopes of NCM neurons, consistent with the hypothesis that adaptation and memory formation in NCM are modulated by the local activation of D1Rs.

1.4 Discussion

In this study, we show that dopamine D1 receptors (D1R) are involved in the auditory association process and modulate excitatory/inhibitory balance in the secondary association pallium (NCM) of a songbird. We show that D1R protein is prevalent in NCM neurons (especially in aromatase-, GABA- and parvalbumin-positive populations). These D1R+ neurons specifically express the plasticity marker EGR1 after classical conditioning. Furthermore, D1R activation regulates NCM circuit excitability in vitro and alters responses and adaptation to song playback in vivo. Together, these pieces of evidence are consistent with the hypothesis that dopamine is an important modulator of complex sensory function and a possible substrate for reinforcement learning.

Dopamine signaling modulates auditory association learning in mammalian primary auditory cortex (Bao et al., 2001; Reichenbach et al., 2015; Schicknick et al., 2012, 2008).
these studies, dopamine activation was paired with simple stimuli such as pure and frequency-modulated tones. However, there is limited work on the role of dopamine in modulating complex auditory signal processing, in high-order cortical/pallial structures. In humans, systemic dopamine-enhancing treatments enhanced auditory language learning (Breitenstein et al., 2004; Knecht et al., 2004). Here, we show that dopamine signaling in high-order sensory pallium can modulate complex auditory signal encoding by shifting inhibitory-excitatory balance.

A distinct feature of NCM among the songbird auditory forebrain nuclei is the dense labeling for somatic aromatase, an enzyme that mediates conversion of testosterone into estradiol (E2) (Saldanha et al., 2000). Many studies describe the important roles of NCM E2 in social interactions and modulation of neural responses (Krentzel et al., 2019; Lampen et al., 2017; Maney et al., 2006; Remage-Healey et al., 2010, 2008; Remage-Healey and Joshi, 2012). We recently showed that blocking aromatase locally in NCM slows auditory association learning in an operant task (Macedo-Lima and Remage-Healey, 2020), which led to the hypothesis that dopamine interacts with E2 signaling in NCM to support association learning. In the present study, we provide anatomical evidence to support this hypothesis (Fig. 1C), suggesting that E2 and dopamine signaling act in tandem to modulate learning and memory in the songbird auditory pallium. We hypothesize that this co-modulatory signaling could apply to other brain circuits that domicile both aromatase and dopamine receptors, such as human auditory cortex (Yague et al., 2006).

GABAergic inhibition fundamentally controls neuronal responses to sounds (Razak and Fuzessery, 2009; Wang et al., 2002) and is required for associative learning in mammal auditory cortex (Letzkus et al., 2011). Dopamine receptors are prevalent in mammalian cortical inhibitory interneurons, especially in parvalbumin-positive (PV+) neurons (Le Moine and Gaspar, 1998).
PV+ cortical interneurons are generally characterized by their fast action potentials and high sustained firing frequency and play a central role in regulating microcircuit activity states and learning processes in hippocampus and cortex (Cardin, 2018; Donato et al., 2013). Here, we show that 55% of the parvalbumin-expressing neurons in NCM also express D1R. We hypothesize that many of the physiological effects we observed can be attributed to this subpopulation.

D1Rs are generally assumed to increase circuit excitability through Gαs-protein coupling (Beaulieu and Gainetdinov, 2011). In our in vitro experiments, D1R agonist SKF-38393 caused a seemingly counterintuitive reduction in the amplitude of GABAergic and glutamatergic currents. However, D1R-mediated depression is well documented in the mammalian nucleus accumbens for both GABAergic and glutamatergic (especially NMDA-mediated) synapses and is attributed to presynaptic plasticity (Pennartz et al., 1992; Zhang et al., 2014). In our in vitro proposed model (Fig. 9A), we suggest two scenarios for the reduction of GABA release: 1) D1Rs are predominantly mediating an increase in GABAergic tone by neurons upstream to those providing input to the recorded neuron, therefore reducing GABA release downstream or 2) D1Rs are acting directly on neurons providing input to the recorded neuron causing a direct reduction in GABA release. For the reduction of glutamatergic current amplitude, we suggest a presynaptic mechanism, perhaps explained by a depletion in presynaptic glutamate stores, resulting from the increased firing (Staley et al., 1998), regulation of glutamate production (Sherman and Mott, 1985), or conversion of glutamate into GABA by glutamine decarboxylase (Neary et al., 1972) by D1Rs. Alternatively, SKF-38393 could be acting directly on the recorded neuron (postsynaptically) in NCM to result in amplitude reductions. Further experiments with miniature events or with synaptic stimulation could help clarify these questions.
In our in vivo recordings, the characteristics of the different cell types led us to formulate hypotheses based on mammalian cortical neurons. Our NS1 neurons are highly reminiscent of mammalian fast-spiking PV+ interneurons, exhibiting short and symmetrical action potentials, high firing, short latency to respond and low stimulus selectivity (Atallah et al., 2012; Tremblay et al., 2016). Because of the narrower waveform, lower firing rate, and longer onset latencies, we suggest that NS2 resemble late-spiking interneurons, such as somatostatin+ or VIP+ interneurons (Tremblay et al., 2016). Regarding our broader waveforms, BS1 has a higher spontaneous firing rate than BS2 and BS2 is not remarkably sensitive to SKF-38393. We found ~33% of NCM neurons do not express either GABA or D1Rs (Fig. 3A). Therefore, it is possible that BS2 cells are part of a circuit in which D1R-signaling does not participate to produce strong effects on the variables analyzed in this study.

We propose three models (Fig. 9B) for the effects we observed in vivo. In our “connected model 1”, we suggest that the D1R activation might be increasing the tonic firing of GABAergic neurons upstream to NS1 cells, thus inhibiting them and disinhibiting BS1 cells. This model resembles a disinhibitory circuitry discovered in mammalian cortex for auditory associative learning, in which learning activates layer 1 inhibitory interneurons, which in turn inhibit layer 2/3 PV+ interneurons, thus disinhibiting pyramidal neurons. These layer 1 neurons are activated by cholinergic signaling (Letzkus et al., 2011), and are known to be 5HT3a+/VIP– interneurons (Tremblay et al., 2016). Alternatively, our “connected model 2” depicts a single synapse and inhibitory effects of SKF-38393 on NS1 cells. Finally, our “disconnected model” summarizes our findings in each cell type and depicts isolated effects of the D1R agonist. Future experiments involving genetic targeting of specific neuronal subtypes could clarify these circuit properties.
Acetylcholine has been shown to affect stimulus-specific adaptation (SSA) in mammalian auditory cortex and inferior colliculus (Ayala and Malmierca, 2015; Metherato and Weinberger, 1989). However, to our knowledge, dopamine modulation of SSA in vertebrate auditory cortex/pallium has not been explored. In mammalian auditory cortex, D1R modulation improves signal detection and association learning (Happel et al., 2014; Schicknick et al., 2012). Likewise, in humans, systemic dopaminergic treatments improve auditory language associative learning (Breitenstein et al., 2004; Knecht et al., 2004). These studies suggest it is plausible that dopamine could be affecting SSA. In songbird NCM, SSA has been shown to parallel familiarity with sounds, such that novel sounds will produce more negative slopes (i.e. higher SSA) than familiar, previously adapted sounds (Chew et al., 1996). In fact, after successful behavioral association learning, learned sounds produce less SSA than novel sounds (Bell et al., 2015). Here, we provide evidence that D1Rs are involved in this process, such that pharmacological D1R activation disrupts SSA in NCM neurons.

We note that the origins of dopaminergic projection to NCM are still an open question in the field. There are 8 major subpallial dopaminergic nuclei, which are fairly well conserved across vertebrates (Reiner et al., 1998). Preliminary reports suggest the caudal ventral tegmental area projects to NCM (Barr et al., 2019; Yanagihara et al., 2019), which, if confirmed, would be an exciting avenue for studying auditory reward prediction learning. Other reports suggest that the locus coerules (LC) projects to NCM to provide norepinephrinergic input (Chen et al., 2016; Ikeda et al., 2015). Norepinephrine is a precursor to dopamine, and LC neurons are known to release dopamine in addition to NE throughout the cortex (Devoto et al., 2005). Dopamine released by the LC onto dorsal hippocampus is involved in spatial learning and memory in rodents, independently of NE release (Kempadoo et al., 2016; Takeuchi et al., 2016).
Furthermore, future studies should clarify through neuronal tract tracing which specific nuclei provide dopaminergic inputs to NCM and whether the effects observed in this study can be mimicked by dopamine release from such nuclei.

In conclusion, we show that D1R signaling shifts the excitatory-inhibitory balance in songbird pallium to modulate mechanisms involved in auditory learning and key components of auditory response, circuitry, and plasticity. We propose that D1Rs are important mediators of learning and memory in the avian sensory pallium and this mechanism could be a common feature among vertebrates.

1.5 References


doi:10.1523/JNEUROSCI.2767-18.2019


arousing effects. *Neuropsychopharmacology* 29:1704–1714. doi:10.1038/sj.npp.1300464


Gadagkar V, Puzerey PA, Chen R, Baird-Daniel E, Farhang AR, Goldberg JH. 2016. Dopamine
neurons encode performance error in singing birds. *Science* (80-) **354**.
doi:10.1126/science.aah6837


doi:10.1126/science.aax0287


Krentzel AA, Ikeda MZ, Oliver TJ, Koroveshi E, Remage-Healey L. 2019. Acute neuroestrogen blockade attenuates song-induced immediate early gene expression in auditory regions of


Lüscher C, Malenka RC. 2012. NMDA receptor-dependent long-term potentiation and long-term


Schmidt MF, Ding L. 2014. Achieving perfection through variability: the basal ganglia helped me do it! *Neuron* **82**:6–8. doi:10.1016/j.neuron.2014.03.010


### 1.6 Figure legends

**Figure 1** – Aromatase and D1 receptor (D1R) proteins are coexpressed by caudomedial nidopallium (NCM) neurons. (A) Songbird auditory circuits. Lighter-to-darker colors illustrate the trajectory of auditory (blue) and motor (red) information. (B) Specificity confirmation of the anti-D1R antibody used here. Preincubation with blocking peptide virtually eliminated cellular staining. (C) Triple immunofluorescence staining for D1R, aromatase and tyrosine hydroxylase (TH; enzyme participating in dopamine synthesis). Magenta and cyan arrows indicate single-labeled D1R and aromatase+ neurons, respectively. White arrows indicate double-labeling. (D) TH fibers are frequently found in close association with double-labeled neurons. Note that fibers tightly envelop double-labeled somata (arrows). Av: nucleus avalanche; ARO: aromatase; CMM: caudomedial mesopallium; D1R: dopamine D1 receptor; HVC: HVC (proper name); Ov: nucleus
ovoidallis; NCM: caudomedial nidopallium; NIf: nucleus interfacialis; RA: robust nucleus of the arcopallium; TH: tyrosine hydroxylase.

**Figure 2** – Early-growth response 1 (EGR1) protein expression in NCM neurons after song stimulation alone or auditory classical conditioning with pure tone and social reinforcement. (A-B) Song stimulation increases EGR1 expression in D1R+ and ARO+ across areas and hemispheres and D1R+/ARO+ neurons in the vNCM across hemispheres as compared to Silence. (C) Classical conditioning behavioral apparatus and (D) timeline. After behavioral manipulations, a 50 min lights-off period followed, then animals were quickly perfused and brains were processed for EGR1, D1R and aromatase (ARO) labeling (A,E). (E) D1R+/EGR1+ neurons are higher in the Paired than in the Unpaired group. Note that in left ventral NCM the Paired group mean is within the range of the Song Reference group. (F) Screen-time predicts EGR1 expression in D1R+ neurons and is higher in Paired than in Unpaired groups. (G) GLM/ANOVA $\chi^2$ results for the frequency of D1R+/EGR1+ neurons after classical conditioning paradigm within each hemisphere*area combination, using Condition and Scree-time as predictors. Significant effects were found for Condition and not Screen-time in the L-vNCM, and of Screen-time and not Condition in the L-dNCM (Paired > Unpaired in both). *p<0.05, **p<0.01, ***p<0.001; L: left; R: right; NCM: caudomedial nidopallium; dNCM: dorsal NCM, vNCM: ventral NCM. Marker colors represent data from individual subjects within each experiment.

**Figure 3** – Dopamine D1 receptor (D1R)-containing neurons are predominantly GABA+. (A) Triple immunofluorescence stain for D1R, GABA and parvalbumin (PV). Magenta, cyan and yellow arrows indicate single-labeled D1R, GABA and PV+ neurons, respectively. White arrows indicate triple-labeling. (B) The majority of PV+ neurons also express GABA and D1Rs.
Figure 4 – Dopamine D1 receptor (D1R) activation with SKF-38393 reduces spontaneous GABAergic (sIPSC) and glutamatergic (sEPSC) current amplitudes but increases sEPSC frequency. (A) Representative neuron in NCM visualized during (left; Alexa-488) and after recording (Neurobiotin-Streptavidin-488 stain; aromatase stain to confirm NCM location). (B) Representative sIPSCs recording from the same NCM neuron, before and during 10 μM SKF-38393 infusion (left panel). Amplitude of sIPSCs (middle panel) was significantly different from rundown (aCSF) and reduced from baseline by SKF (4-7 min). Frequency (right panel) was also reduced but did not differ from rundown experiments (5-7 min). Waveforms show representative sIPSC (mean±SEM during 1 min) before (PRE), on the last minute of (SKF) and on the last minute after (POST) drug infusion. (C) Representative AMPA/NMDA/Kainate sEPSCs recording from the same NCM neuron, before and during 10 μM SKF-38393 infusion (left panel). Amplitude of sEPSCs (middle panel) was significantly reduced by 10 and 50 μM (no dose difference) SKF (6-7 min). In addition, frequency (right panel) was increased by SKF (2, 5 min). In these experiments, rundown were performed before treatment and did not significantly change. Waveforms show representative sEPSC (mean±SEM during 1 min) before (PRE), on the last minute of (SKF) and on the last minute after (POST) drug infusion. *p<0.05. Downward current traces in B and C represent depolarizing current deflections.

Figure 5 – Cell type separation based on waveform measurements in in vivo recordings. (A) Unsupervised hierarchical clustering results (gap statistic) using peak-to-peak ratio vs duration (top left); separation into 4 clusters best explained the variance in the data. Traces illustrate all waveforms in our sample by cell type (middle left). Shaded lines are the average waveforms of each unit in our sample; opaque lines and shading are the grand average±SEM of those. The pie chart represents the proportions of units within each cluster in our sample (bottom left).
correlation plot shows the 4 distinct clusters and the kernel-density estimations of the distributions along single-axis with mean±SEM on top. (B) Representative PSTHs and adaptation slopes of NS1 (left) and BS1 (right) cell types. Representative examples in B are highlighted in all graphs according to cluster type. (C) Physiology parameters of the different cluster types. From left to right: spontaneous and stimulus firing rate, z-score, adaptation slope, latency to respond to song and percentage of songs responded to. Tukey’s post-hoc test results are shown at the bottom. Significant differences (p<0.05) are assigned different letters. *p<0.05, **p<0.01, ***p<0.001. Highlighted datapoints in A,C depict the two representative cells in B.

**Figure 6** – RetroDrive recordings, i.e. acute multichannel electrophysiology microdrive coupled to a microdialysis probe for drug delivery through retrodialysis. (A) RetroDrives were constructed using a custom electrode interface board, a 32-channel Omnetics connector, a CMA11 (Harvard Apparatus) microdialysis probe, and 32x 12.5-μm-diameter NiCr wires spun into tetrodes. Scale bar: 5 mm. (B) Tetrodes were ~0.2 mm distant from the probe and cut to extend to approximately half the length of the probe. Scale bar: 1 mm. (C) Timeline for drug infusions and playbacks starting after auditory site localization within NCM. (D) Electrode track confirmation. Before insertion in the brain, electrodes were dipped in DiI-594 (6.25% in 200-proof ethanol) for posthumous site confirmation.

**Figure 7** – Effects of D1R activation *in vivo* on firing characteristics of NCM neurons. (A) Representative PSTHs of an NS1 (top) and a BS1 (bottom) cell in response to SKF. Note the change in stimulus and spontaneous firing in NS1 and BS1, respectively, due to SKF-38393 (0.2 mM). Representative cells are highlighted in all plots according to their cell type. (B) Spontaneous firing rate. D1R agonist reduced firing of NS1, but increased firing of BS1 cells (trend in full GLM model, but significant in restricted model). (C) Stimulus firing rate. D1R
agonist reduced firing of NS1 cells. (D) Correlation between spontaneous and stimulus %-change during SKF infusion accompanied by kernel-density estimations of the distributions along single-axis with mean±SEM on top. Regression line across all data points is shown (solid line). Identity line is shown as a dotted line. The regression slope 95% confidence interval (shaded area) is less than and does not encompass 1, indicating that spontaneous firing changes are higher than stimulus firing changes. (E) D1R agonist reduced Z-score, (F) increased latency and (G) decreased responsiveness of NCM cells. Relevant statistical effects are highlighted on top of each plot. Post-hoc Tukey test results are displayed on the plots when interaction was significant. **p<0.01; *p<0.05; *p<0.1.

**Figure 8** – D1R activation disrupts neuronal stimulus-specific adaptation *in vivo*. (A) Representative PSTHs from a single NS1 highlighting a decrease in adaptation slope to a new stimulus due to SKF-38393 (0.2 mM) treatment. (B) The slope through the normalized firing rate of trials 6-25 averaged by cell type. (C) Adaptation slopes quantifications show a reduction in the slope due to SKF-38393. Relevant statistical effect is displayed on the top of the plot. *p<0.05.

**Figure 9** – Network connectivity models hypothesized from current findings. SKF-38393 effects are shown by the large arrows. (A) Models for *in vitro* results with sIPSCs (GABAergic) and sEPSCs (glutamatergic). GABAergic model 1 hypothesizes that D1R agonist binds preferentially to a GABAergic neuron upstream from the ones providing input to the recorded neuron. Binding of D1R agonist increases its GABA tone resulting in a reduction of GABA release immediately upstream of the recorded neuron. GABAergic model 2 suggests D1R effects directly cause a reduction in GABA release by the upstream neuron. Glutamatergic effects can be explained by a direct effect of D1R agonist on a glutamatergic neuron directly upstream from the recorded neuron, causing a reduction (perhaps a depletion) of glutamate release but an increase in firing...
rate. (B) Models for in vivo results. Connected models only depict cell types with changes in firing rate. Model one hypothesizes that D1R agonist binds to and excites a GABAergic neuron upstream of NS1 cells, which results in a reduction of NS1 firing and consequent disinhibition of BS1 cells. Model 2 offers a direct effect of D1R agonist on NS1 cells, resulting in a reduction in their firing, which results in disinhibition of BS1 cells. In the disconnected model, the effects observed here are due to D1R’s binding to all cell types individually (except BS2, in which no effects were observed). sI/EPSC: spontaneous inhibitory/excitatory postsynaptic current.
<table>
<thead>
<tr>
<th>Antibody</th>
<th>Type</th>
<th>Host</th>
<th>Dilution</th>
<th>Company</th>
</tr>
</thead>
<tbody>
<tr>
<td>anti-D1DR</td>
<td>Polyclonal</td>
<td>Guinea pig</td>
<td>1:100</td>
<td>Alomone Labs</td>
</tr>
<tr>
<td>D1DR peptide</td>
<td>-</td>
<td>Rat</td>
<td>1:10</td>
<td>Alomone Labs</td>
</tr>
<tr>
<td>anti-Aromatase</td>
<td>Polyclonal</td>
<td>Rabbit</td>
<td>1:2000</td>
<td>Azac</td>
</tr>
<tr>
<td>anti-Aromatase</td>
<td>Polyclonal</td>
<td>Sheep</td>
<td>1:2000</td>
<td>Bethyl Labs</td>
</tr>
<tr>
<td>anti-EGR1</td>
<td>Polyclonal</td>
<td>Rabbit</td>
<td>1:2000</td>
<td>Santa Cruz</td>
</tr>
<tr>
<td>anti-Tyrosine Hydroxylase</td>
<td>Monoclonal</td>
<td>Mouse</td>
<td>1:2000</td>
<td>Immunostar</td>
</tr>
<tr>
<td>anti-GABA</td>
<td>Polyclonal</td>
<td>Rabbit</td>
<td>1:1000</td>
<td>Sigma</td>
</tr>
<tr>
<td>anti-Parvalbumin</td>
<td>Monoclonal</td>
<td>Mouse</td>
<td>1:10000</td>
<td>Millipore</td>
</tr>
</tbody>
</table>
FIGURE 6

A

B

C

D