Prepubertal ovariectomy alters dorsomedial striatum indirect pathway neuron excitability and explore/exploit balance in female mice

Kristen Delevich[1,2*], Christopher D. Hall[1], and Linda Wilbrecht[1,2*]
1 Department of Psychology, University of California, Berkeley, CA 94720, USA
2 Helen Wills Neuroscience Institute, University of California, Berkeley, CA 94720, USA

*co-corresponding author

Linda Wilbrecht, PhD
Department of Psychology
University of California, Berkeley
16 Barker Hall
Berkeley, CA 94720-1650

Tel: +1.510.600.3560
wilbrecht@berkeley.edu

Kristen Delevich, PhD
Department of Integrative Physiology and Neuroscience
College of Veterinary Medicine
Washington State University
P.O. Box 647620 Pullman, WA 99164

Tel: +1.509-335-4829
kristen.delevich@wsu.edu

Abstract:

Decision-making circuits are modulated across life stages (e.g. juvenile, adolescent, or adult)—as well as on the shorter timescale of reproductive cycles in females—to meet changing environmental and physiological demands. Ovarian hormonal modulation of relevant neural circuits is a potential mechanism by which behavioral flexibility is regulated in females. Here we examined the influence of prepubertal ovariectomy (pOVX) versus sham surgery on performance in an odor-based multiple choice reversal task. We observed that pOVX females made different types of errors during reversal learning compared to sham surgery controls. Using reinforcement learning models fit to trial-by-trial behavior, we found that pOVX females exhibited lower inverse temperature parameter (β) compared to sham females. These findings suggest that OVX females solve the reversal task using a more exploratory choice policy, whereas sham females use a more exploitative policy prioritizing estimated high value options. To seek a neural correlate of this behavioral difference, we performed whole-cell patch clamp recordings within the dorsomedial striatum (DMS), a region implicated in regulating action selection and explore/exploit choice policy. We found that the intrinsic excitability of dopamine receptor type 2 (D2R) expressing indirect pathway spiny projection neurons (iSPNs) was significantly higher in pOVX females compared to both unmanipulated and sham surgery females. Finally, to test whether mimicking this increase in iSPN excitability could recapitulate the pattern of reversal task behavior observed in pOVX females, we chemogenetically activated DMS D2R(+) neurons within intact female mice. We found that chemogenetic activation increased exploratory choice during reversal, similar to the pattern we observed in pOVX females. Together, these data suggest that pubertal status may influence explore/exploit balance in females via the modulation of iSPN intrinsic excitability within the DMS.

71 **Introduction:**
72
73 As animals interact with their environment in pursuit of rewards in the form of food, water,
74 mates etc., they learn from trial and error to guide their future choices. This process involves
75 learning from positive and negative feedback and also, importantly, deciding how learned
76 information should influence choice, referred to as choice policy. Reinforcement learning (RL)
77 models (Sutton and Barto, 1998) have provided a useful framework for understanding and
78 quantifying aspects of trial-and-error learning, including choice policy. A classic RL problem that
79 hinges on choice policy is the explore/exploit tradeoff. If an animal (or any agent for that matter)
80 adopts an exploit policy, it will consistently select the highest estimated value option, but may
81 miss out on better alternative options. On the other hand, if an animal favors a more exploratory
82 choice policy, characterized by less value-dependent, more stochastic choice behavior, it may
83 discover new and better options more readily (Sutton and Barto, 1998; Daw et al., 2006).
84 Importantly, the optimal balance of exploration and exploitation may depend on the statistics of
85 the environment and/or the needs of the animal as defined by its particular physiological or
86 developmental state (Cohen et al., 2007; Frank et al., 2009; Humphreys et al., 2015; Addicott et
87 al., 2017; Lenow et al., 2017; Gopnik, 2020). In humans, choice behavior generally becomes less
88 exploratory and more exploitative during the transition from childhood to adulthood (Nussenbaum
89 and Hartley, 2019; Gopnik, 2020; Xia et al., 2020; Eckstein et al., 2021). Natural fluctuations in
90 ovarian hormones across the estrous cycle or exogenous estradiol administration have been shown
91 to regulate aspects of value-based decision making in female rats (Uban et al., 2012; Orsini et al.,
92 2021), including explore/exploit balance (Verharen et al., 2019b). These data suggest that the rise
93 in ovarian hormones at puberty could contribute to the developmental shift in choice policy during
94 adolescence in females. In previous work, we observed that pOVX altered performance in a

95 multiple choice reversal task in adult C57/Bl6 mice (Delevich et al., 2020a). Compared to intact
96 females, pOVX females showed lower ratios of perseverative to regressive errors during reversal
97 learning, but the potential underlying biological processes that contributed to this behavioral effect
98 remained unclear.

99 The DMS is implicated in the regulation of goal-directed action selection (Tai et al., 2012;
100 Nonomura et al., 2018; Matamales et al., 2020; Peak et al., 2020) and choice policy (Collins and
101 Frank, 2014), and recent work suggests that enhancing the activity of D2R(+) SPNs in the dorsal
102 striatum biases choice behavior to be more exploratory (Lee et al., 2015; Delevich et al., 2020b)
103 but see (Verharen et al., 2019a). While nuclear estrogen receptors are notably absent from the
104 dorsal striatum in adulthood (Krentzel et al., 2021), extranuclear estrogen receptors (ERα, ERβ,
105 and GPER1) localize to SPNs, glia, and the presynaptic terminals of striatal GABAergic and
106 cholinergic interneurons of adult female rats (Almey et al., 2012). At the neuronal level, estrous
107 cycle has been shown to regulate the intrinsic excitability of SPNs located within the rodent
108 striatum (Proano et al., 2018; Alonso-Caraballo and Ferrario, 2019). Studies examining the
109 influence of estrous cycle on SPN physiology have been primarily performed in rats, where SPN
110 cell types were not distinguished, but see (Tansey et al., 1983). Taken together, these findings raise
111 the question of whether pubertal status influences choice strategies employed by females by
112 modulating striatal SPN physiology. Here we focused on D2R(+) SPNs of the indirect pathway
113 (iSPNs) within the DMS, whose activity we hypothesized regulates explore/exploit balance in
114 decision making based on theoretical predictions (Collins and Frank, 2014; Dunovan and
115 Verstynen, 2016) plus genetic (Beeler et al., 2010; Kwak et al., 2014) and pharmacological (Lee
116 et al., 2015; McCoy et al., 2019) evidence.

117 In the current study, after analyzing raw behavioral data, we applied reinforcement learning
118 modeling to examine how pOVX influenced learning and choice policy processes underlying
119 performance in the odor-based reversal learning task. We next examined the influence of pOVX
120 on the intrinsic excitability of genetically identified D2R(+) SPNs within the DMS of adult female
121 mice. Finally, we chemogenetically activated D2R(+) neurons within the DMS of female mice
122 during reversal learning and applied our RL model to determine whether this manipulation
123 recapitulated the reversal learning strategy employed by pOVX females. We found that compared
124 to intact adult females, pOVX females exhibited a more exploratory choice strategy during reversal
125 learning as evidenced by a lower explore/exploit inverse temperature parameter, β. In addition,
126 D2R(+) SPN intrinsic excitability was increased in pOVX females compared to sham females.
127 Finally, chemogenetic activation of D2R(+) SPNs within the DMS promoted a more exploratory
128 choice strategy during reversal learning in intact female mice, resembling pOVX female behavior.
129 Together, these data suggest that pubertal status influences the choice strategy female mice employ
130 via the modulation of D2R(+) SPN activity.

131 **Materials & Methods:**

132 ***Animals:***

133 Female C57BL/6NCR (Charles River), Drd2-eGFP BAC (GENSAT), and D2-Cre ER43
134 (MMRC) mice were bred in-house. Drd2-eGFP BAC and D2-Cre ER43 mice were bred onto the
135 C57BL/6NCR background for at least 5 generations. All mice were weaned on postnatal day (P)21
136 and housed in groups of 2–3 same-sex siblings on a 12:12 hr reversed light:dark cycle (lights on
137 at 2200 h). All behavioral tests were conducted during the dark phase. For all experiments, mice
138 were randomly assigned to experimental groups and sample sizes were based on previously
139 conducted experiments (e.g. Delevich et al. 2020a,b). Each behavioral experiment was conducted

140 once, and no animal was tested on multiple occasions. All procedures were approved by the
141 Animal Care and Use Committee of the University of California, Berkeley and conformed to
142 principles outlined by the NIH Guide for the Care and Use of Laboratory Animals.

143 ***Prepubertal Ovariectomy:***

144 Prepubertal ovariectomy was performed as previously described (Delevich et al., 2020a).
145 To eliminate ovarian hormone exposure during and after puberty, ovariectomies were performed
146 before puberty onset at P25. Prior to ovariectomy (OVX) surgery, all female mice were visually
147 inspected to confirm that vaginal opening had not occurred. Prior to surgery, mice were injected
148 with 0.05 mg/kg buprenorphine and 10 mg/kg meloxicam subcutaneously and were anesthetized
149 with 1–2% isoflurane during surgery. The incision area was shaved and scrubbed with ethanol and
150 betadine. Ophthalmic ointment was placed over the eyes to prevent drying. A 1 cm incision was
151 made with a scalpel in the lower abdomen across the midline to access the abdominal cavity. The
152 ovaries were clamped off from the uterine horn, with locking forceps and ligated with sterile
153 sutures. After ligation, the ovaries were excised with a scalpel. The muscle and skin layers were
154 sutured, and wound clips were placed over the incision for 7–10 days to allow the incision to heal.
155 An additional injection of 10 mg/kg meloxicam was given 24 and 48 h after surgery. Sham control
156 surgeries were performed in which fat pads were visualized but the ovaries were not clamped,
157 ligated, or excised. Female littermates were randomly assigned to sham or pOVX groups. Mice
158 were allowed to recover on a heating pad until ambulatory and were post-surgically monitored for
159 7–10 days to check for normal weight gain and signs of discomfort/distress. Mice were co-housed
160 with 1-2 siblings who received the same surgical treatment. To confirm the success of prepubertal
161 ovariectomies, necropsy was performed on a subset of adult sham and ovariectomized mice to

162 confirm that the uteri of pOVX mice were underdeveloped compared to age-matched sham females
163 (data not shown).

164 ***4 choice odor-based reversal task:***

165 Sham or pOVX mice were tested in an odor-based reversal task that has previously been
166 described in detail (Johnson and Wilbrecht, 2011; Johnson et al., 2016) as young adults (P60-P70).
167 The task is designed such that only the odor cue is predictive of reward, while spatial and
168 egocentric information is irrelevant. Briefly, mice were food restricted to ~85% body weight by
169 the Discrimination phase. Mice were habituated to the testing arena on day 1, they were taught to
170 dig for a honey nut cheerio reward in a pot filled with unscented wood shavings on day 2,
171 underwent a 4 choice odor Discrimination on day 3, and finally, were tested on Recall of the
172 previously learned odor-reward association, which was immediately followed by a Reversal phase
173 on day 4. During the Discrimination phase of the task, mice learned to discriminate among four
174 pots with different scented wood shavings (anise, clove, litsea, and thyme). All four pots were
175 sham-baited with cheerio (under wire mesh at bottom) but only one pot was rewarded (anise). The
176 pots of scented shavings were placed in each corner of an acrylic arena (12”, 12”, 9”) which was
177 divided into four quadrants. Mice were placed in a cylinder in the center of the arena, and a trial
178 started when the cylinder was lifted. Mice were then free to explore the arena and indicate their
179 choice by making a bi-manual dig in one of the four pots of wood shavings. The cylinder was
180 lowered as soon as a choice was made. If the choice was incorrect, the trial was terminated and the
181 mouse was gently encouraged back into the start cylinder. Trials in which no choice was made
182 within 3 minutes were considered omissions. If mice omitted for two consecutive trials, they
183 received a reminder: a baited pot of unscented wood shavings was placed in the center cylinder
184 and mice dug for the “free” reward. Mice were disqualified if they committed four pairs of

185 omissions. The location of the four odor scented pots was shuffled on each trial, and criterion was
186 met when the mouse completed 8 out of 10 consecutive trials correctly. 24 hours after completing
187 Discrimination, mice were tested for Recall of the initial odor Discrimination to criterion, after
188 which, mice immediately proceeded to the Reversal phase in which the previously rewarded odor
189 (anise) was no longer rewarded, and a previously unrewarded odor (clove) was now rewarded.
190 During the Reversal phase, Odor 4 (thyme) was replaced by a novel odor (eucalyptus) that was
191 unrewarded. Again, mice were run until they reached a criterion of 8 out of 10 consecutive correct
192 trials.

193 ***4 choice odor-based reversal task analysis:***

194 To compare reversal task performance across groups, trials to criterion and errors (incorrect
195 choices) were compared for each phase of the task (Discrimination, Recall, and Reversal).
196 Omission trials did not count towards trials to criterion. In addition, for the Reversal phase we
197 separated errors in which mice chose the odor that was rewarded during Discrimination (Odor 1)
198 into two types: 1) perseverative errors occurred when Odor 1 was chosen prior to the first correct
199 trial and 2) regressive errors occurred when Odor 1 was chosen after the first correct trial during
200 the Reversal phase. To compare the relative proportion of these error types within mice, we
201 calculated Reversal error bias as (perseverative – regressive errors)/(perseverative + regressive
202 errors). Therefore, a value > 1 indicates a bias for perseverative errors whereas a value < 1 indicates
203 a bias for regressive errors. Finally, we examined how quickly mice accumulated rewards after the
204 first correct trial during the Reversal phase by aligning trial histories to the first correct trial and
205 summing rewarded trials across the subsequent 8 trials. Data were fit by linear regression for each
206 group and the slope of the lines compared to determine whether groups significantly differed in
207 their rate of reward accumulation. Behavioral data from 14 of the 16 pOVX females and 10 of the

208 15 sham females presented here were included in a previously published study examining sex
209 differences of prepubertal gonadectomy on approach-avoidance behaviors, but latent decision
210 variables were not examined (Delevich et al., 2020a).

211 ***Reinforcement learning modeling of 4 choice odor-based reversal task***

212 We modeled Discrimination and Reversal phase behavior using a reinforcement learning
213 model driven by an iterative error-based rule (Rescorla and Wagner, 1972; Sutton and Barto,
214 1998). The model uses a prediction error (δ) to update the value (V) of each odor stimulus, where
215 δ is the difference between the experienced feedback (λ) and the current expected value (r= 100
216 for rewarded, r= 0 for unrewarded) scaled by a learning rate parameter (α), with 0<α<1:

217 $$V(t+1)_{choice} = V(t)_{choice} + \alpha \times \delta(t)$$

218 $$\delta(t) = r(t) - V(t)_{choice}$$

219 $$r(t) \begin{cases} 100 \; if \; rewarded \; choice \\ else \; 0 \end{cases}$$

220

221 Because mice exhibit innate preferences for odors, we set initial odor values to fixed
222 parameters [v1,v2,v3,v4] for all mice tested by calculating the probability of choosing each odor
223 during the first 4 trials of Discrimination × 100 (see Johnson et al. 2016). These initial odor values
224 were calculated separately for mice included in Figure 1 and Figure 4 (see data source files or
225 analysis code for more details). To model trial-by-trial choice probabilities, the stimulus values
226 were transformed using a softmax function to compute choice probabilities based on estimated
227 odor values, $V(O)_i$. The inverse temperature parameter (β), which we refer to in the text as the
228 explore/exploit parameter, determined the stochasticity of the choices:

229 $$\sigma(V(O))_i = \frac{e^{\beta V(O)_i}}{\sum_{i=1}^{n} e^{\beta V(O)_i}}$$

230 For RL modeling, trial histories from Discrimination and Recall phases were concatenated
231 to create one Discrimination phase trial history. We compared the alternative models using AIC
232 (Watanabe, 2010) and found that the best fit model included phase-specific (non-zero) α and β
233 parameters; all RL model comparisons for pOVX and sham females are presented in Table S1 as
234 well as source data files and analysis code. To assess model performance, trial-by-trial behavioral
235 data was recovered using the best fit parameters for each animal, and average recovered choices
236 to criterion for Discrimination and Reversal phases (100 simulations/animal) were plotted against
237 the actual choices to criterion for each animal.

238 ***Stereotaxic Virus Injection***

239 Female D2-Cre mice (6-8 weeks) were deeply anesthetized with 5% isoflurane (vol/vol) in
240 oxygen and placed into a stereotactic frame (Kopf Instruments; Tujunga, CA) upon a heating pad.
241 Anesthesia was maintained at 1-2% isoflurane during surgery. An incision was made along the
242 midline of the scalp and small burr holes were drilled over each injection site. Virus was delivered
243 via microinjection using a Nanoject II injector (Drummond Scientific Company; Broomall, PA).
244 Injection coordinates for DMS were (in mm from bregma): 0.90 anterior, +/-1.4 lateral, and -3.0
245 from surface of the brain. Adeno-associated viruses (AAVs) were produced by Addgene viral
246 service and had titers of $>10^{12}$ genome copies per mL. For chemogenetic manipulations, mice were
247 bilaterally injected with 0.5 uL of rAAV8-hsyn-DIO-mCherry (N=9) rAAV8-hsyn-DIO-hM3Dq-
248 mCherry (N=6), or rAAV8-hsyn-DIO-hM4Di-mCherry (N=5). Mice were given subcutaneous
249 injections of meloxicam (10 mg/kg) during surgery and 24 and 48 hours after surgery. Mice were
250 group-housed before and after surgery and 4-6 weeks were allowed for viral expression before
251 behavioral training or electrophysiology experiments.

252 ***Drugs***

253 Clozapine-N-Oxide was generously provided by the NIMH Chemical Synthesis and Drug
254 Supply Program (NIMH C-929). CNO was made fresh each day and dissolved in DMSO (0.5%
255 final concentration) and diluted to 0.1 mg/mL in 0.9% saline USP.

256 ***Electrophysiology***

257 Mice were deeply anesthetized with an overdose of ketamine/xylazine solution and
258 perfused transcardially with ice-cold cutting solution containing (in mM): 110 choline-Cl, 2.5 KCl,
259 7 $MgCl_2$, 0.5 $CaCl_2$, 25 $NaHCO_3$, 11.6 Na-ascorbate, 3 Na-pyruvate, 1.25 NaH_2PO_4, and 25 D-
260 glucose, and bubbled in 95% O_2/5% CO_2. 300 µm thick coronal sections were cut in ice-cold
261 cutting solution before being transferred to ACSF containing (in mM): 120 NaCl, 2.5 KCl, 1.3
262 $MgCl_2$, 2.5 $CaCl_2$, 26.2 $NaHCO_3$, 1 NaH_2PO_4 and 11 Glucose. Slices were bubbled with 95%
263 O_2/ 5% CO_2 in a 37°C bath for 30 min, and allowed to recover for 30 min at room temperature
264 before recording. All recordings were made using a Multiclamp 700B amplifier and were not
265 corrected for liquid junction potential. The bath was heated to 32°C for all recordings. Data were
266 digitized at 20 kHz and filtered at 1 or 3 kHz using a Digidata 1440 A system with pClamp 10.2
267 software (Molecular Devices, Sunnyvale, CA, USA). Only cells with access resistance of <25 MΩ
268 were retained for analysis. Cells were discarded if parameters changed more than 20%. Data were
269 analyzed using pClamp or R (RStudio 0.99.879; R Foundation for Statistical Computing, Vienna,
270 AT).

271 Whole-cell current clamp recordings were performed using a potassium gluconate-based
272 intracellular solution (in mM): 140 K Gluconate, 5 KCl, 10 HEPES, 0.2 EGTA, 2 $MgCl_2$, 4
273 MgATP, 0.3 Na_2GTP, and 10 Na_2-Phosphocreatine. Alexa Fluor 594 (40 µM) was added to the
274 internal solution to enable morphological confirmation of SPN identify following recording. In
275 order to block NMDA and AMPA-mediated currents, 5 µM AP5 and 25 µM NBQX were added

276 to the ACSF, respectively for intrinsic excitability data in Figure 2. For all recordings, cells were
277 allowed to stabilize for 2 min after break in and prior to any current injection. For current clamp
278 recordings to test the effect of CNO in Gq-DREADD- expressing vs. mCherry-expressing D2R(+)
279 neurons, baseline input-output curves were collected before 5 minute wash-on of 10 μM CNO.

280 ***Histology***

281 Mice were transcardially perfused with PBS followed by 4% PFA in PBS. Following 24h
282 postfixation, coronal brain slices (75 μm) were sectioned using a vibratome (VT100S Leica
283 Biosystems; Buffalo Grove, IL). To confirm viral targeting, we performed a standard
284 immunohistochemical procedure using a primary antibody against red fluorescence protein (RFP)
285 (rabbit, Rockland 600-401-379; 1:1000) to enhance the mCherry signal expressed in mice
286 transduced with rAAV8-hSyn-DIO-DREADD-mCherry or rAAV8-hSyn-DIO-mCherry. Sections
287 were counterstained with DAPI (Life Technologies; Carlsbad, CA). Images were acquired with a
288 Zeiss Axio Scan.Z1 epifluorescence microscope (Molecular Imaging Center, UC Berkeley) at 10x
289 magnification and viewed using FIJI (ImageJ). Anatomical regions were identified according to
290 the Mouse Brain in Stereotaxic Coordinates by Franklin and Paxinos and the Allen Institute Mouse
291 Brain Atlas.

292 ***Statistics and Data Analysis***

293 For comparisons between 2 groups, a t-test was used when data were normally distributed, and
294 Welch's correction was applied when variance was unequal. The D'Agostino & Pearson test was
295 used to test for normality. For experiments in which 3 groups were compared, a one-way ANOVA
296 or Kruskal Wallis test when not normally distributed was performed, followed by two-tailed
297 uncorrected Fisher's LSD or Dunn's test, respectively, for pairwise comparisons. Two-way
298 ANOVA was performed when two independent variables were examined (e.g. treatment and error

type), followed by uncorrected Fisher's LSD (two-tailed) for pairwise comparisons. Post-hoc comparisons were not corrected, due to the limited number of planned comparisons. Throughout the paper, p=0.05 was used as the criterion for a significant statistical difference unless noted otherwise. Data are expressed as mean ± SEM unless noted otherwise.

Data availability

All data generated or analyzed during this study are included in the manuscript and supporting files. Source data files have been provided for all experiments reported in this manuscript in an online repository at https://doi.org/10.6084/m9.figshare.14783628.v1. Analysis code is available at https://github.com/kdelevich/4choice_RLmodeling.

Results:

Prepubertal ovariectomy affects reversal learning by promoting exploratory choice policy

We performed sham surgery or pOVX on female C57/Bl6 mice at postnatal day 25 (P25), prior to puberty onset, and trained them in an odor-based reversal task between P60-70 (Fig. 1A). The odor-based reversal task consisted of two main phases: 1) a Discrimination phase during which mice learned through trial and error that one of four scented pots of wood shavings contained a buried food reward and 2) a Reversal phase in which the odor-reward contingency was reversed (Fig. 1B). Sham females were not staged for estrous cycle, and both groups performed similarly in the Recall phase (Supplementary Fig. 1). When comparing Discrimination and Reversal, there was a significant effect of task phase but not treatment on trials to reach criterion [task phase: $F(1,29)= 6.31$, $p= 0.018$; treatment: $F(1,29)= 0.11$, $p= 0.74$; task phase × treatment: $F(1,29)= 0.05$, $p= 0.83$] (Fig. 1C).

Next, we more closely examined the types of errors that mice made during Reversal. Error types included those made to the previously rewarded odor, which we divided into 2 subtypes:

322 perseverative (errors made before the first correct trial) and regressive (errors made after first
323 correct trial). Perseverative errors reflect a tendency to stick to a previously learned rule, whereas
324 regressive errors reflect a failure to acquire or maintain the new rule. There was a significant
325 interaction between error type and treatment group [$F(5,145)= 2.79$, $p=0.02$] (Fig. 1D). Post hoc
326 analyses revealed that pOVX females made significantly more regressive errors compared to sham
327 females ($p= 0.03$ uncorrected Fisher's LSD). We next examined the pattern of perseverative and
328 regressive errors made by individual mice. Sham females exhibited a significantly higher ratio of
329 perseverative to regressive errors (Reversal error bias) compared to pOVX females (sham vs.
330 pOVX females: $t(29)=2.12$, $p= 0.04$) (Figure 1E). Finally, we observed that sham females
331 accumulated rewards at a significantly higher rate after the first rewarded trial compared to pOVX
332 females during Reversal but not Discrimination (Figure 1F). These data suggest that sham females
333 and pOVX females reach criterion in the reversal task using different trial-by-trial strategies.

334 We next turned to computational modeling to assess if differences observed in the Reversal
335 phase between sham and pOVX females arise from a difference in odor value updating, a
336 difference in choice policy, or a combination of both. To do so, we fit trial-by-trial behavioral data
337 with RL models and used the maximum log likelihood to determine the parameters that best fit
338 each animal's behavior. The best fit model included phase-specific parameters for the learning rate
339 α and the explore/exploit inverse temperature parameter β (Fig. 1G) (see Supplementary Table 1
340 for alternate model comparison). We found that there was a significant interaction between task
341 phase and treatment for the explore/exploit parameter β [task phase $\times$ treatment: $F(1,29) = 7.101$,
342 $p= 0.013$]. In sham and pOVX female mice, the explore/exploit parameter was significantly higher
343 during the Reversal phase compared to Discrimination phase (sham Reversal vs. Discrimination:
344 $p<0.0001$; pOVX Reversal vs. Discrimination $p= 0.011$ uncorrected Fisher's LSD) and Reversal

phase explore/exploit parameter was significantly lower in pOVX vs. sham females (pOVX vs. sham: p= 0.014 uncorrected Fisher's LSD) consistent with pOVX females employing a more exploratory choice policy compared to sham females (Fig. 1H).

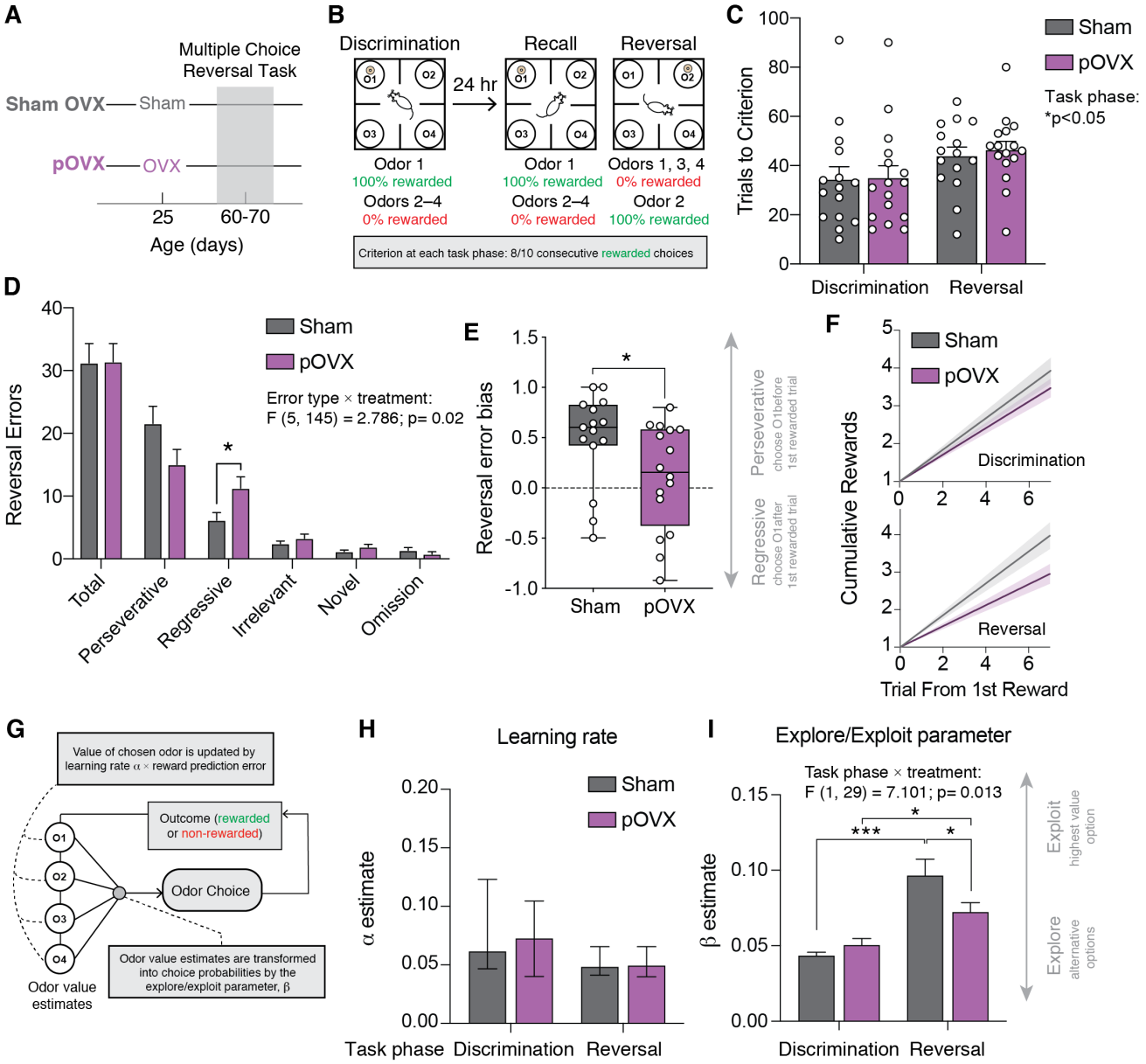


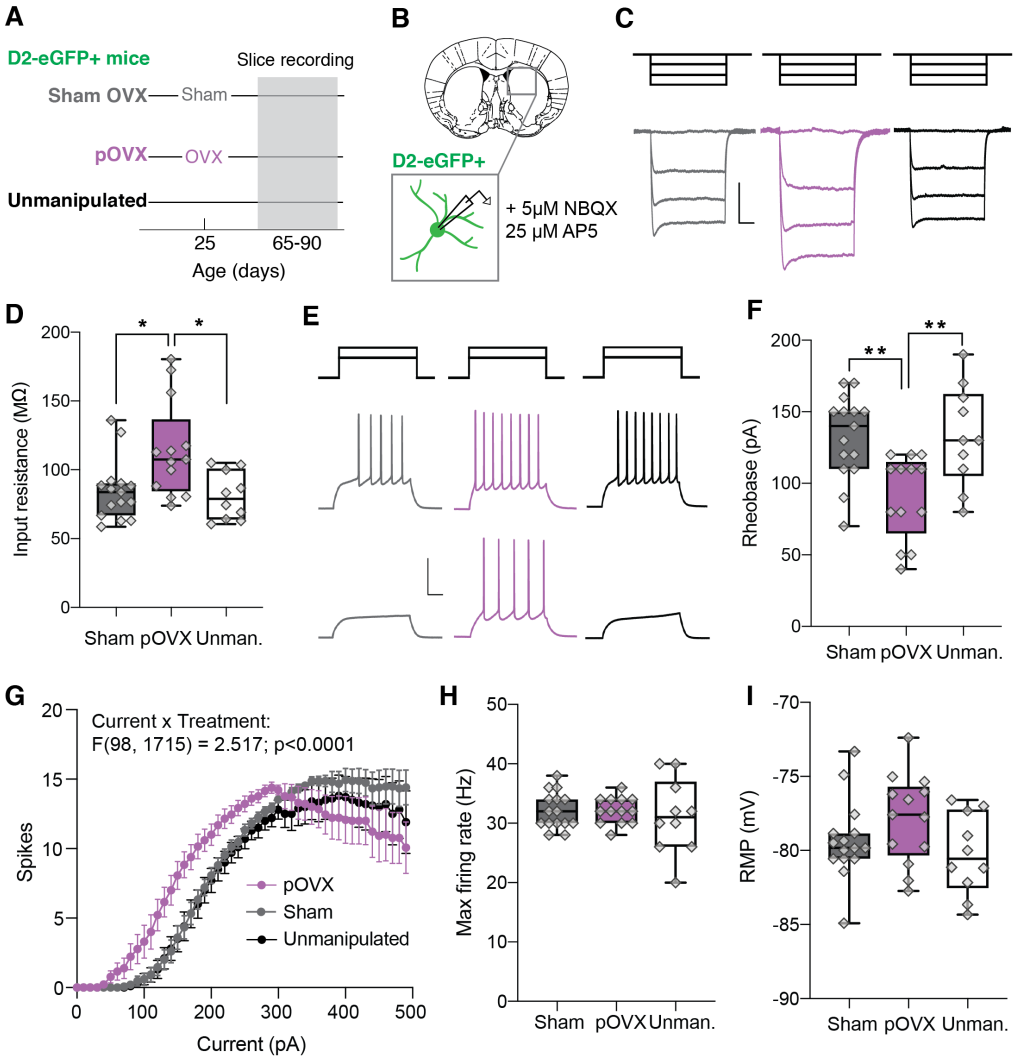
Figure 1. Prepubertal OVX is associated with more exploratory reversal learning strategy in female mice. (A) Female C57/Bl6 mice underwent OVX or sham surgery at P25 and were trained in the multiple choice reversal task in adulthood (P60-70). **(B)** Mice were trained to a criterion of

352 8/10 correct consecutive choices to Odor 1 during Discrimination. 24 hours later they were tested
353 for Recall of the previous day's rule before immediately advancing to a Reversal phase during
354 which Odor 2, rather than Odor 1, was rewarded. Reversal criterion was reached when mice made
355 8/10 correct consecutive choices to Odor 2. **(C)** There was a main effect of task phase on trials to
356 criterion but no effect of treatment (Two-way RM ANOVA main effect of task phase $F(1, 29) =$
357 6.30, $p<0.05$). **(D)** There was a significant effect of treatment and error type on the number of
358 reversal errors (Two-way RM ANOVA treatment × error type interaction: $F(5, 145) = 2.79$,
359 $p<0.05$). pOVX females made significantly more regressive errors compared to sham females
360 (11.25 ± 1.8 vs. 6.13 ± 1.2, $p<0.05$, uncorrected Fisher's LSD). **(E)** pOVX females had a
361 significantly lower Reversal error bias (perseverative – regressive errors)/(perseverative +
362 regressive errors) compared to sham females (0.11 ± 0.13 vs. 0.49 ± 0.12, $p<0.05$, unpaired t-test).
363 **(F)** Sham females accumulated rewards after the first correct Reversal trial faster than pOVX
364 females (best fit line with 95% C.I. plotted). **(G)** RL model applied to odor-based multiple choice
365 reversal task. Schematic based on (Verharen et al., 2019b). **(H)** Best-fit α learning rate estimates
366 did not significantly differ by task phase or treatment. **(I)** There was a significant interaction
367 between task phase and treatment group on the best-fit explore/exploit β parameter (Two-way
368 ANOVA task phase × treatment interaction: $F(1,29)= 7.101$, $p<0.05$). Post-hoc comparisons
369 revealed that β parameter estimates were significantly higher during Reversal compared to the
370 Discrimination phase for sham ($p<0.0001$, uncorrected Fisher's LSD) and pOVX females ($p<0.05$,
371 uncorrected Fisher's LSD). In addition, Reversal phase β parameter estimates were significantly
372 lower in pOVX females compared to sham ($p<0.05$, uncorrected Fisher's LSD). Data in (H) plotted
373 as median ± IQR.

374 **Prepubertal OVX is associated with increased intrinsic excitability of D2R(+) SPNs**

375 The DMS has been implicated in action selection and determining choice policy, and
376 previous studies have found evidence that estrous cycle modulates the intrinsic excitability of
377 striatal SPNs. Furthermore, several lines of evidence suggest that D2R(+) iSPNs: 1) are modulated
378 by ovarian hormones (Le Saux and Di Paolo, 2005; Le Saux et al., 2006; Krentzel et al., 2019) and
379 2) can influence explore/exploit balance during decision making (Kwak et al., 2014; Lee et al.,
380 2015; Delevich et al., 2020b). We therefore investigated whether changes in the intrinsic
381 excitability of D2R(+) SPNs within DMS may contribute to sham vs. pOVX differences in choice
382 policy during reversal learning. We performed whole-cell current clamp recordings of visually
383 identified eGFP+ and neurons within the DMS of adult D2-eGFP transgenic female mice who
384 underwent pOVX or sham surgery and unmanipulated female mice in the presence of the
385 excitatory synaptic blockers NBQX and AP5 (Fig. 2A-C). AlexaFluor-594 was included in the
386 internal solution, and all cells included in analysis were confirmed to have spinous morphology.
387 We found a main effect of treatment on D2-eGFP(+) SPN input resistance [main effect of
388 treatment: $H= 8.76$, $p= 0.0125$] with pOVX females exhibiting higher input resistance compared
389 to sham and unmanipulated females (pOVX vs. sham $p=0.013$; pOVX vs. unman. $p=0.009$,
390 uncorrected Dunn's test) (Fig. 2D). When we injected a series of positive current steps (Fig. 2E),
391 we found that the minimum amount of current necessary to trigger an action potential (rheobase)
392 was significantly lower in pOVX females compared to sham and unmanipulated females (pOVX
393 vs. sham $p=0.001$; pOVX vs. unman. $p= 0.003$, uncorrected Fisher's LSD) (Fig. 2F). In addition,
394 there was a significant interaction between treatment and current on spike output [$F(98, 1715)=$
395 2.517, $p<0.0001$] (Fig. 2G). While input-output curves were shifted leftward in pOVX compared
396 to sham and unmanipulated females, there was no significant effect of treatment on maximum
397 firing rate [$F(2, 18.68)= 0.10$, $p= 0.90$] (Fig. 2H). Finally, resting membrane potential (RMP) did

398 not differ across treatments [F(2, 35)= 2.172, p= 0.129] (Fig. 2I). These data indicate that D2R(+)
399 SPNs within the DMS are more intrinsically excitable in pOVX females compared to unstaged
400 sham and unmanipulated female mice.



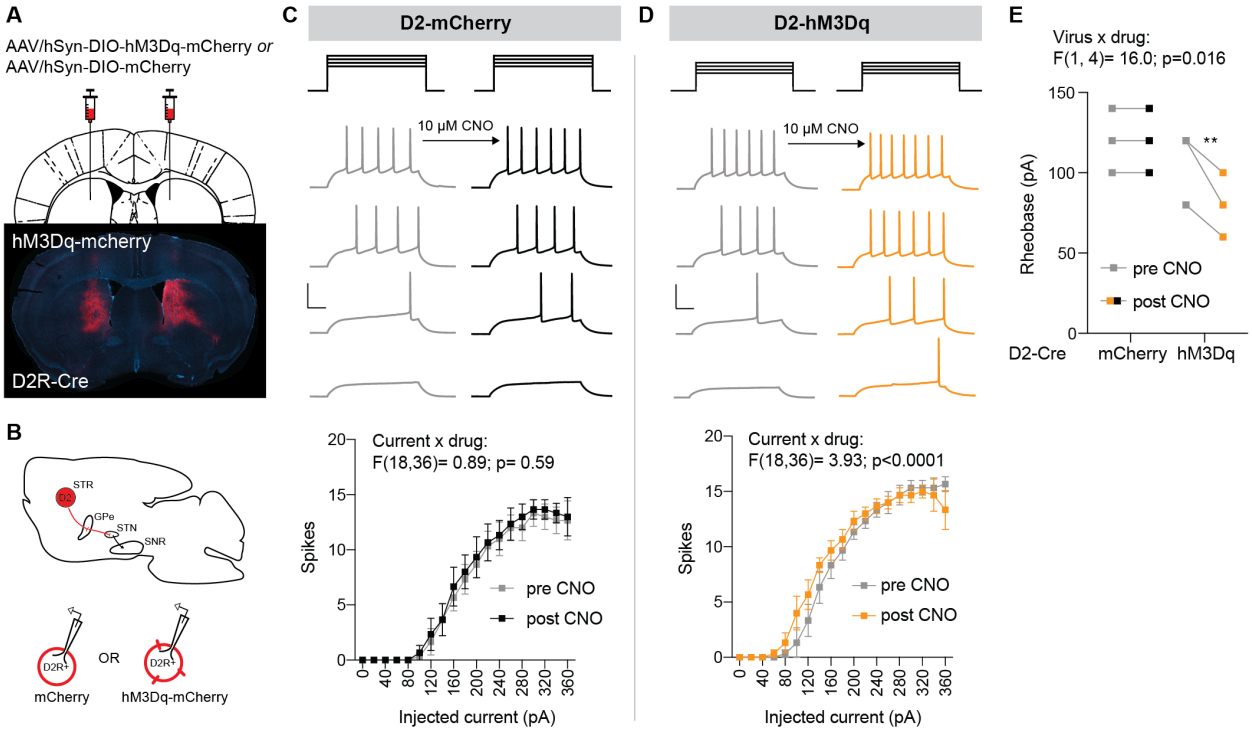
401 **Figure 2. Prepubertal OVX increases D2R(+) SPN intrinsic excitability in the presence of**
402 **synaptic blockers. (A)** At P25 D2-eGFP(+) female mice underwent sham or pOVX surgery, while
403 a third group of female D2-eGFP(+) mice received no surgery. **(B)** Whole-cell current clamp
404 recordings were made from visually identified D2-eGFP(+) SPNs within the DMS from all groups
405 in adulthood (P65-90). **(C)** Representative responses to negative current steps (-150, -100, -50, 0
406 pA) in D2R(+) SPNs from sham, pOVX, and unmanipulated females. Scale bar: 100 ms, 5 mV.

407 **(D)** D2-eGFP(+) SPNs in pOVX female mice had higher input resistance compared to sham and
408 unmanipulated females. **(E)** Representative responses to positive current steps (120, 180 pA) in
409 D2R(+) SPNs from sham, pOVX, and unmanipulated females. Scale bar: 100 ms, 50 mV. **(F)**
410 Decreased rheobase of D2-eGFP(+) SPNs was observed in pOVX compared to sham and
411 unmanipulated females. **(G)** Spike number across sequential depolarizing current steps (10-500
412 pA) for D2-eGFP(+) SPNs. Increased spiking was observed in pOVX compared to sham and
413 unmanipulated females (Two-way RM ANOVA current x treatment interaction: $F(98, 1715)=$
414 2.52, $p<0.0001$). **(H)** No difference in maximum firing rate was observed across treatment groups.
415 **(I)** No difference in RMP was observed across treatment groups. *$p<0.05$, **$p<0.01$; n/N = 15/5,
416 13/5, and 10/3 for sham, pOVX, and unmanipulated mice, respectively.

417 **Chemogenetic activation of D2R(+) SPNs in DMS reduces perseverative bias and promotes**
418 **a more exploratory reversal strategy in female mice**

419 Given that OVX females exhibit a more exploratory reversal strategy and greater intrinsic
420 excitability of D2R(+) SPNs in DMS, we next asked whether experimentally increasing D2R(+)
421 SPN intrinsic excitability would similarly bias intact female mice towards increased exploration
422 during the Reversal phase. Female D2-Cre mice were bilaterally infused with 0.5 μL of Cre-
423 dependent DREADD virus (hM4Di-mCherry or hM3Dq-mCherry) and trained 4-6 weeks later in
424 the 4 choice odor-based reversal learning task (Fig. 3A). Female mice expressing Cre-inducible
425 mCherry were used to control for any effects of surgery, AAV infection, and clozapine-N-oxide
426 (CNO) administration on behavior. To examine how CNO activation of hM3Dq expressed by
427 D2R(+) SPNs in DMS alters their activity, we performed whole-cell current clamp recordings of
428 identified mCherry+ neurons in mice that expressed the excitatory DREADD hM3Dq or mCherry
429 alone (Fig. 3B). Briefly, spike output in response to depolarizing steps (0–360 pA, 20 pA steps)

430 was recorded from visually identified mCherry+ neurons in D2-mCherry or D2-hM3Dq-mCherry
431 expressing SPNs in DMS (Fig. 3A-D). Next, 10 µM CNO was bath-applied for 5 minutes and
432 spike output to the same sequential series of depolarizing current steps was recorded (Fig. 3A-D).
433 There was no significant interaction between current step and drug on spike output in D2-mCherry
434 expressing SPNs [Two-way RM ANOVA, current x drug: F(18,36)=0.89, p=0.59] (Fig. 3C) but
435 there was a significant interaction between current and drug on spike output in D2-hM3Dq-
436 mCherry SPNs [Two-way RM ANOVA, current x drug: F(18,36)=3.93, p=0.0002] (Fig. 3D).
437 Finally, there was a significant interaction between virus and drug on rheobase [Two-way RM
438 ANOVA F(1,4)= 16.0, p=0.016] (Fig. 3E).



439 **Figure 3. CNO increases intrinsic excitability of hM3Dq-expressing iSPNs. (A)** Schematic of
440 injection and representative brain section showing hM3Dq-mCherry expression in the DMS of
441 D2-Cre mouse. **(B)** Schematic of indirect pathway expression (sagittal view) and whole-cell patch-
442 clamp configuration of mCherry+ or hM3Dq-mCherry+ iSPNs in female D2-Cre mice. **(C)** Top

443 panel: representative responses to positive current steps (100, 120, 140, 160 pA) in mCherry+
444 iSPNs before and after CNO wash on. Scale bar: 100 ms, 50 mV. Bottom panel: no significant
445 interaction between current step and CNO treatment on spike output in D2R(+) mCherry-
446 expressing iSPNs. **(D)** Top panel: representative responses to positive current steps (100, 120, 140,
447 160 pA) in hM3Dq-mCherry+ iSPNs before and after CNO wash on. Scale bar: 100 ms, 50 mV.
448 Bottom panel: significant interaction between current step and CNO treatment on spike output in
449 D2R(+) mCherry-expressing iSPNs (Two-way ANOVA current step × drug, $p<0.0001$). **(E)**
450 Summary of CNO wash on effect on rheobase (Two-way ANOVA virus × drug, $p<0.05$).

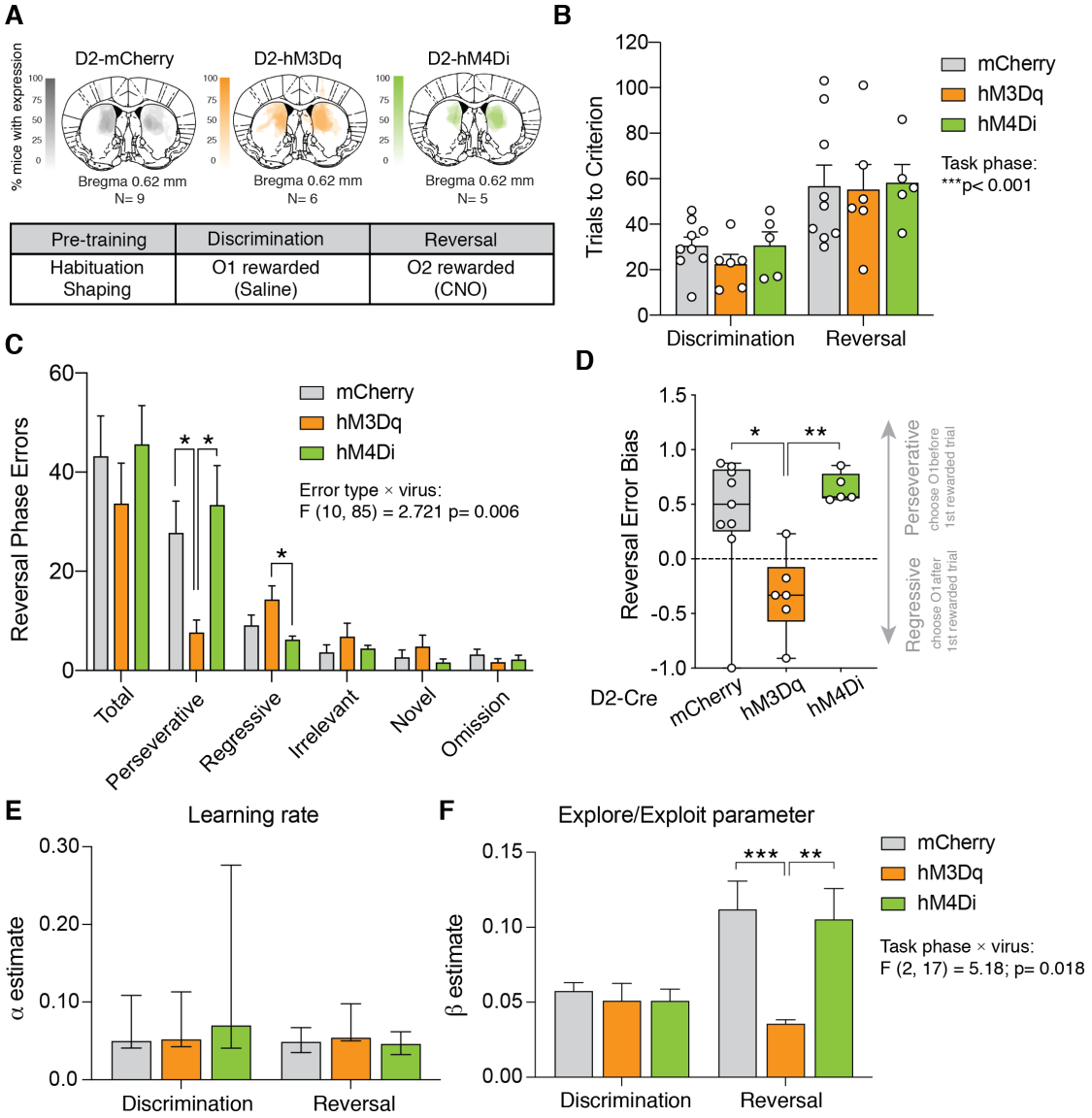
451 Prior to Discrimination training all mice received i.p. injections of saline (Fig. 4A) and
452 learned through trial and error that one of four presented odors indicated the location of a buried
453 food reward. Mice completed the Discrimination task phase when they selected the rewarded odor
454 (Odor 1) on 8/10 consecutive trials. Twenty-four hours later, all groups were administered CNO
455 (1.0 mg/kg, i.p.) and tested for their recall of discrimination learning followed immediately by a
456 Reversal phase in which Odor 1 was no longer rewarded and Odor 2 became rewarded. There was
457 a significant effect of task phase on trials to criterion [Reversal vs. Discrimination; $F(1,17)= 16.58$,
458 $p= 0.0008$] but no significant effect of virus [$F(2,17) = 0.37$, $p= 0.69$] or interaction between virus
459 and task phase [$F(2,17)= 0.09$, $p=0.92$] (Fig. 4B). While there was no significant effect of
460 chemogenetic manipulation on Reversal phase trials to criterion, we found a significant interaction
461 between virus and error type during Reversal [$F(10,85)= 2.721$, $p= 0.006$] (Fig. 4C) that was absent
462 during Discrimination when mice were on saline (Supplementary Figure 2). D2-hM3Dq mice
463 made significantly fewer perseverative errors compared to D2-mCherry ($p= 0.015$, uncorrected
464 Dunn's test) and D2-hM4Di groups ($p= 0.028$, uncorrected Fisher's LSD) and made significantly
465 more regressive errors compared to D2-hM4Di mice ($p= 0.03$, uncorrected Fisher's LSD) (Fig.

466 4C). We next examined whether chemogenetic manipulation of D2R(+) neurons in the DMS
467 altered Reversal error bias within mice. There was a significant effect of virus on Reversal error
468 bias (H= 9.06, p= 0.005 Kruskal-Wallis test), with D2-hM3Dq mice having a significantly lower
469 Reversal error bias compared to D2-mCherry (p=0.019, uncorrected Dunn's test) and D2-hM4Di
470 groups (p= 0.005, uncorrected Dunn's test) (Fig. 4D), consistent with a greater tendency to make
471 regressive errors compared to perseverative errors. This data suggests that chemogenetic activation
472 of D2R(+) neurons in the DMS produced a pattern of reversal phase choice behavior that was
473 similar to the effect seen in pOVX mice.

474 Finally, we applied RL modeling to determine whether similar changes in decision-making
475 parameters might explain the pattern of reversal behavior we observed in pOVX female mice and
476 D2-hM3Dq female mice. Fitting the same RL model (task phase-specific α and β parameters; see
477 Methods) we found that there was no significant interaction between virus and task phase on
478 learning rate α [F(2,17)= 0.45, p=0.64] (Fig. 4E), but there was a significant interaction between
479 virus and task phase for the explore/exploit parameter β [F(2,17)= 5.18, p=0.018] (Fig. 4F). The
480 Reversal phase explore/exploit parameter was significantly lower in D2-hM3Dq mice compared
481 to D2-mCherry (p=0.0002, uncorrected Fisher's LSD) and D2-hM4Di (p= 0.002, uncorrected
482 Fisher's LSD) (Fig. 4F). These data suggest that chemogenetic activation of D2R(+) neurons
483 within DMS biases choice strategy in female mice to be more exploratory during reversal learning.
484 Moreover, chemogenetic activation of D2R(+) neurons within the DMS produced behavior in
485 female mice that mimicked the behavioral pattern seen in OVX females, including a reduction in
486 Reversal error bias during reversal learning and a reduced explore/exploit β parameter consistent
487 with a less exploitative, more exploratory choice policy. Taken together with evidence that D2R(+)
488 iSPNs within the DMS are more intrinsically excitable in pOVX compared to sham females, these

489 data support a model whereby pOVX biases reversal learning strategy towards exploration by

490 modulating iSPN intrinsic excitability within DMS (Fig. 5).



491 **Figure 4. Chemogenetic activation of D2R(+) neurons in DMS promotes a more exploratory**

492 **reversal strategy in intact female mice. (A)** Top panel: schematic illustrating injection site and

493 viral spread female D2-Cre DIO-mCherry (N=9), DIO-hM3Dq (N=6), and DIO-hM4Di (N=5)

494 mice. Bottom panel: summary of behavioral training. **(B)** There was a main effect of task phase

495 but no effect of virus on trials to criterion (Two-way RM ANOVA main effect of task phase:

496 $F(2,17)= 16.58$, $p<0.001$). **(C)** There was a significant interaction between error type and virus on
497 reversal errors (Two-way RM ANOVA error type × manipulation interaction: $F(10,85)= 2.72$,
498 $p<0.01$) with D2-hM3Dq mice making fewer perseverative errors ($p<0.05$, uncorrected Fisher's
499 LSD) compared to D2-mCherry and D2-hM4di mice and more regressive errors compared to D2-
500 hM4di mice ($p<0.05$, uncorrected Fisher's LSD). **(D)** There was a main effect of virus on Reversal
501 error bias ($W=9.06$, $p<0.01$ Kruskal-Wallis test), with D2-hM3Dq mice showing reduced bias for
502 perseverative errors compared to D2-mCherry ($p<0.05$, uncorrected Dunn's test) and D2-hM4Di
503 mice ($p<0.01$, uncorrected Dunn's test). **(E)** Best-fit α learning rate did not significantly differ by
504 task phase or virus. **(F)** There was a significant interaction between task phase and treatment group
505 on the best-fit explore/exploit parameter β (Two-way ANOVA task phase x treatment interaction:
506 $F(2,17)= 5.18$, $p<0.05$). Post-hoc comparisons revealed that β parameter estimates were
507 significantly higher during Reversal compared to Discrimination phase for D2-mCherry mice
508 ($p<0.01$, uncorrected Fisher's LSD) and D2-hM4Di mice ($p<0.05$, uncorrected Fisher's LSD) but
509 not D2-hM3Dq mice ($p=0.41$, uncorrected Fisher's LSD). In addition, Reversal phase β parameter
510 estimates were significantly lower in D2-hM3Dq mice compared to D2-mCherry ($p<0.001$,
511 uncorrected Fisher's LSD) and D2-hM4Di mice ($p<0.01$, uncorrected Fisher's LSD). *$p<0.05$,
512 **$p<0.01$, ***$p<0.001$. Data in (E) plotted as median ± IQR.
513
514

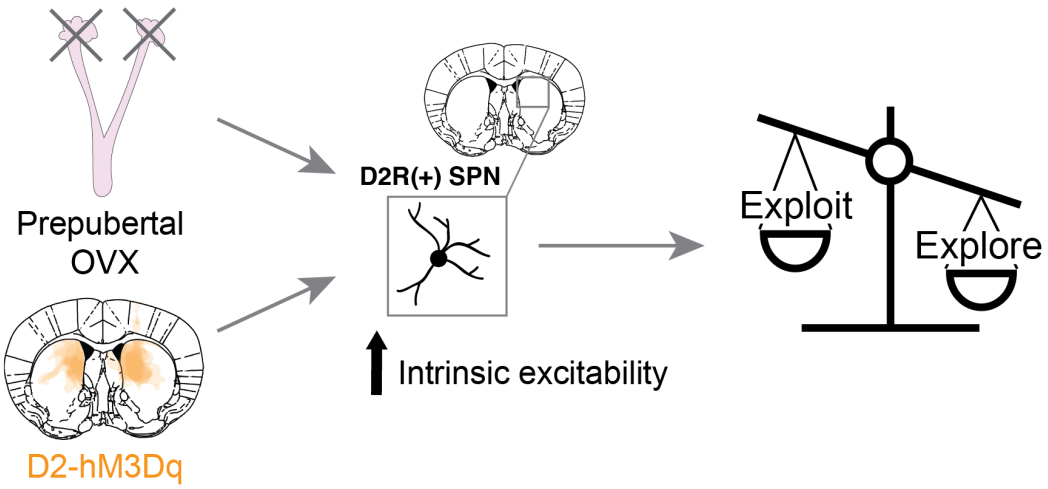


Figure 5. Results summary. Both pOVX and chemogenetic activation of D2R(+) neurons within the DMS are associated with increased intrinsic excitability of iSPNs. In turn, both manipulations are associated with less perseverative, more exploratory choice strategy during reversal learning. While indirect, the convergent behavioral effects of pOVX and chemogenetic activation of DMS D2R(+) neurons suggest that the increased intrinsic excitability of iSPNs in the DMS of pOVX mice could contribute to the altered reversal strategy observed. Future experiments should perform *in vivo* recording of D2R(+) SPNs and/or chemogenetic manipulation experiments in pOVX females to probe the relationship between altered D2R(+) SPN intrinsic properties and reversal learning strategy on a trial-by-trial basis.

Discussion:

We found that pOVX altered how female mice solved a reversal learning task. Using RL models fit to trial-by-trial behavioral data, we found that pOVX mice exhibited a more exploratory choice policy during reversal learning than sham controls, captured by a lower inverse temperature β parameter. This difference in exploratory choice behavior was accompanied by increased intrinsic excitability of D2R(+) iSPNs in the DMS, a region that is implicated in regulating action selection and choice policy. We then sought to mimic this effect using chemogenetics. We

534 demonstrated that chemogenetic activation of D2R(+) neurons *in vitro* similarly enhanced iSPN
535 intrinsic excitability in slices from female brains. In addition, activation of DMS D2R(+) neurons
536 *in vivo* decreased the ratio of perseverative to regressive errors and promoted exploratory choice
537 captured by a lower inverse temperature β parameter. Together, these data suggest that two distinct
538 manipulations: pOVX and hM3Dq activation converged on similar behavioral effects through a
539 shared mechanism of enhancing DMS iSPN intrinsic excitability.

540 Our data are consistent with studies that manipulate D2Rs and model choice policy.
541 Germline D2R knockout (Kwak et al., 2014), systemic D2R antagonist administration (Eisenegger
542 et al., 2014), and intrastriatal D2R antagonist infusion (Lee et al., 2015) are each associated with
543 more exploratory choice policy. However, none of these studies could rule out the contribution of
544 presynaptic D2 autoreceptors, which is important given the apparent role of tonic dopamine in
545 modulating explore/exploit balance (Beeler et al., 2010; Humphries et al., 2012; Cinotti et al.,
546 2019), but see (Costa et al., 2014). Our chemogenetic manipulation experiments (which do not
547 infect D2R(+) dopamine axon terminals) clearly demonstrate that activation of D2R(+) neurons
548 within DMS is sufficient to bias performance towards exploration.

549 We speculate that there are two likely circuit mechanisms downstream of D2R(+) iSPNs
550 that may be responsible for promoting exploratory choice policy. The first involves local lateral
551 connections from iSPNs to direct pathway SPNs (dSPNs) and the second involved the interface of
552 the direct and indirect pathways in basal ganglia output centers such as the substantia nigra pars
553 reticulata (SNr). One recent study showed that systemic injection of the D2R antagonist raclopride
554 induced dopamine-dependent transcriptional activation in iSPNs that opposed the activation of
555 dSPNs, suggesting that iSPN to dSPN transmodulation is an important mechanism for behavioral
556 flexibility (Matamales et al., 2020). Therefore, it is possible that elevated iSPN activity, either

557 through pOVX or chemogenetic activation, promotes exploratory choice by dampening the
558 activity of task-relevant ensembles of dSPNs that would normally promote the selection of the
559 highest estimated-value option. Opponent mechanisms between the direct and indirect pathway at
560 convergent downstream targets are also predicted to regulate choice policy (Collins and Frank,
561 2014).

562 Studies have shown that the intrinsic properties of striatal SPNs differ between females and
563 males before puberty (Dorris et al., 2015) and in an estrous cycle-dependent manner after puberty
564 (Proano et al., 2018). Interestingly, Proano et al. found that the intrinsic excitability of accumbal
565 SPNs was significantly higher during diestrus/metestrus, when estradiol and progesterone levels
566 are low, compared to proestrus/estrus (Proano et al., 2018). This may be in keeping with our
567 observation that the intrinsic excitability of D2R(+) SPNs was higher in pOVX females compared
568 to sham females, since pOVX females lack gonadally-produced estradiol and progesterone.
569 Estradiol has also been shown to influence dopamine release and reuptake in the striatum (Calipari
570 et al., 2017), including the dorsal striatum (Becker and Beer, 1986; Becker, 1990). Recent data
571 also suggest that dopamine influences the postnatal maturation of intrinsic excitability of SPN
572 populations within the striatum (Lieberman et al., 2018). Therefore, it is possible that the changes
573 we observed in D2R(+) SPN excitability in pOVX female mice may occur by direct action on
574 SPNs or downstream of hormonal effects on presynaptic dopamine release (Lin et al., 2020).

575 There are several lines of evidence that suggest that ovarian hormones preferentially
576 modulate the activity of D2R(+) iSPNs versus D1R(+) dSPNs. OVX decreases D2 receptor
577 binding in the striatum, and estradiol or treatment with an ERβ agonist counteracts the effect of
578 OVX (Le Saux et al., 2006). The aforementioned treatments did not alter D2R mRNA expression,
579 suggesting that estradiol modulates D2R binding through a mechanism other than transcriptional

580 regulation of D2R expression. Furthermore, OVX reduces the expression of preproenkephalin,
581 which produces the endogenous opioid peptide, enkephalin, which is expressed in iSPNs (Le Saux
582 and Di Paolo, 2005). Again, the effect of OVX on preproenkephalin expression can be
583 counteracted by estradiol administration. Finally, in the nucleus accumbens core, rapid
584 enhancement of mEPSC amplitude by estradiol is inversely correlated with rheobase (Krentzel et
585 al., 2019). Given that D2R(+) SPNs typically display lower rheobase compared to D1R+ SPNs,
586 this suggests that estradiol may exert a greater effect on D2R(+) iSPNs compared to D1R(+)
587 dSPNs. However, it should be noted that the authors did not observe rapid effects of estradiol on
588 mEPSC amplitude within dorsal striatum (Krentzel et al., 2019).

589 There are several limitations to our current study that should be noted. First, we cannot
590 assume that the changes we observed in SPN intrinsic excitability are specific to the dorsomedial
591 region of the striatum or to the D2R(+) SPN cell type. Second, our evidence linking the increased
592 excitability of D2R(+) SPNs in DMS to the more exploratory choice strategy used by pOVX
593 females is correlational. In independent experiments we observed that 1) pOVX promoted
594 exploratory choice strategy during Reversal 2) pOVX is associated with elevated intrinsic
595 excitability of D2R(+) SPNs in DMS and 3) that chemogenetic activation of D2R(+) SPNs in DMS
596 promoted exploratory choice strategy during Reversal. In the future, more direct evidence could
597 be gained by performing manipulation experiments to reduce the activity of DMS D2R(+) SPNs
598 in OVX females, or by recording the activity of these same neurons in pOVX and sham females
599 during behavior. Finally, while we performed OVX prior to puberty onset, we do not know whether
600 the timing of OVX plays an important role in the observed effect on behavior and physiology. We
601 also do not know if and when hormone replacement may rescue the effects of pOVX. Future
602 studies should examine timing effects of OVX and hormone replacement on these outcome

603 measures. Still, in light of these limitations, our data suggest future lines of inquiry into the
604 relationship between puberty, ovarian hormones, SPN physiology, and choice policy in value-
605 based decision making.

606 As yet, we have not identified the mechanism by which pOVX leads to enhanced
607 excitability of D2R(+) SPNs. Ovarian hormones have been shown to regulate dendritic complexity
608 and spine density in cell types in other brain regions (Gould et al., 1990; Woolley et al., 1990;
609 Wallace et al., 2006; Chen et al., 2009; Ye et al., 2019). The dendrites of D2R(+) SPNs are enriched
610 in Kir2 family inward rectifying K+ channels (Uchimura et al., 1989; Nisenbaum and Wilson,
611 1995; Mermelstein et al., 1998; Shen et al., 2007), and a reduction in dendritic length/complexity
612 has been associated with reduced Kir2 expression and increased intrinsic excitability (Cazorla et
613 al., 2012; Sebel et al., 2017). Therefore, it would be informative to compare iSPN Kir2 channel
614 currents and dendritic morphology in sham vs. pOVX females. Finally, it is possible that the
615 increase in intrinsic excitability of the D2R(+) SPNs in pOVX females could represent a
616 homeostatic plasticity mechanism that accompanies a reduction in excitatory synaptic inputs to
617 them, but we did not measure synaptic inputs onto D2R(+) SPNs in this study.

618 There is a growing interest in understanding the mechanisms that underlie sex differences
619 in value-based decision making. A recent study showed that compared to males, female mice
620 employed a more consistent strategy while learning a two-dimensional decision-making task
621 (Chen et al., 2021). This observed tendency for female mice to constrain their decision-space align
622 with the exploitative reversal learning strategy we observed in sham females. Conversely, we
623 observed that pOVX females exhibited a markedly more exploratory reversal learning strategy,
624 sticking less to the previously rewarded odor choice and committing more regressive errors
625 compared to sham females. These findings suggest that ovarian hormones contribute to the female-

biased choice strategies utilized during value-based decision making. While previous studies have separately provided evidence that ovarian hormones regulate the intrinsic excitability of SPNs and aspects of value-based decision making, we show for the first time that pOVX alters explore/exploit balance of choice strategy while also increasing the intrinsic excitability of D2R(+) SPNs in the DMS. These data suggest that pubertal status may influence explore/exploit balance via the modulation of SPN intrinsic excitability within the DMS and highlight a role for ovarian hormones in establishing sex-specific decision-making strategies in adulthood. These data can inform the basic science of decision making and the study of the many psychiatric disorders that emerge after puberty and also show sex differences in their prevalence or manifestation.

Acknowledgments

We thank Yuting Zhang and Kenechukwu Okwuosa for technical assistance with mouse behavior testing. We thank Benjamin Hoshal for contributing to analysis code. We thank Dr. Helen Bateup for feedback on the manuscript and Dr. Anne Collins and Wilbrecht lab members for helpful discussion.

Declaration of interests

The authors declare that there are not conflicts of interest.

References:

Addicott MA, Pearson JM, Sweitzer MM, Barack DL, Platt ML (2017) A Primer on Foraging and the Explore/Exploit Trade-Off for Psychiatry Research. Neuropsychopharmacol 42:1931-1939.

Almey A, Filardo EJ, Milner TA, Brake WG (2012) Estrogen Receptors Are Found in Glia and at Extranuclear Neuronal Sites in the Dorsal Striatum of Female Rats: Evidence for Cholinergic But Not Dopaminergic Colocalization. Endocrinology 153:5373-5383.

Alonso-Caraballo Y, Ferrario CR (2019) Effects of the estrous cycle and ovarian hormones on cue-triggered motivation and intrinsic excitability of medium spiny neurons in the Nucleus Accumbens core of female rats. Horm Behav 116.

Becker JB (1990) Direct Effect of 17-Beta-Estradiol on Striatum - Sex-Differences in Dopamine Release. Synapse 5:157-164.

Becker JB, Beer ME (1986) The Influence of Estrogen on Nigrostriatal Dopamine Activity - Behavioral and Neurochemical Evidence for Both Pre-Naptic and Postsynaptic Components. Behav Brain Res 19:27-33.

Beeler JA, Daw N, Frazier CRM, Zhuang XX (2010) Tonic dopamine modulates exploitation of reward learning. Front Behav Neurosci 4.

Calipari ES, Juarez B, Morel C, Walker DM, Cahill ME, Ribeiro E, Roman-Ortiz C, Ramakrishnan C, Deisseroth K, Han MH, Nestler EJ (2017) Dopaminergic dynamics underlying sex-specific cocaine reward. Nat Commun 8.

Cazorla M, Shegda M, Ramesh B, Harrison NL, Kellendonk C (2012) Striatal D2 Receptors Regulate Dendritic Morphology of Medium Spiny Neurons via Kir2 Channels. J Neurosci 32:2398-2409.

Chen CS, Ebitz RB, Bindas SR, Redish AD, Hayden BY, Grissom NM (2021) Divergent Strategies for Learning in Males and Females. Curr Biol 31:39-+.

Chen JR, Yan YT, Wang TJ, Chen LJ, Wang YJ, Tseng GF (2009) Gonadal Hormones Modulate the Dendritic Spine Densities of Primary Cortical Pyramidal Neurons in Adult Female Rat. Cereb Cortex 19:2719-2727.

Cinotti F, Fresno V, Aklil N, Coutureau E, Girard B, Marchand AR, Khamassi M (2019) Dopamine blockade impairs the exploration-exploitation trade-off in rats. Sci Rep-Uk 9.

Cohen JD, McClure SM, Yu AJ (2007) Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. Philos T R Soc B 362:933-942.

Collins AG, Frank MJ (2014) Opponent actor learning (OpAL): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. Psychological review 121:337-366.

Costa VD, Tran VL, Turchi J, Averbeck BB (2014) Dopamine Modulates Novelty Seeking Behavior During Decision Making. Behav Neurosci 128:556-566.

Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006) Cortical substrates for exploratory decisions in humans. Nature 441:876-879.

Delevich K, Hall CD, Piekarski D, Zhang YT, Wilbrecht L (2020a) Prepubertal gonadectomy reveals sex differences in approach-avoidance behavior in adult mice. Horm Behav 118.

Delevich K, Hoshal BD, Zhang Y, Vedula S, Collins AGE, Wilbrecht L (2020b) Activation But Not Inhibition of the Indirect Pathway Disrupts Choice Suppression in a Freely Moving, Multiple Choice Foraging Task. In: Available at SSRN.

701 Dorris DM, Cao JY, Willett JA, Hauser CA, Meitzen J (2015) Intrinsic excitability varies by sex
702 in prepubertal striatal medium spiny neurons. J Neurophysiol 113:720-729.
703 Dunovan K, Verstynen T (2016) Believer-Skeptic Meets Actor-Critic: Rethinking the Role of
704 Basal Ganglia Pathways during Decision-Making and Reinforcement Learning. Front
705 Neurosci-Switz 10.
706 Eckstein MK, Master SL, Dahl RE, Wilbrecht L, Collins AGE (2021) The Unique Advantage of
707 Adolescents in Probabilistic Reversal: Reinforcement Learning and Bayesian Inference
708 Provide Adequate and Complementary Models. bioRxiv:2020.2007.2004.187971.
709 Eisenegger C, Naef M, Linssen A, Clark L, Gandamaneni PK, Muller U, Robbins TW (2014)
710 Role of Dopamine D2 Receptors in Human Reinforcement Learning.
711 Neuropsychopharmacol 39:2366-2375.
712 Frank MJ, Doll BB, Oas-Terpstra J, Moreno F (2009) Prefrontal and striatal dopaminergic genes
713 predict individual differences in exploration and exploitation. Nat Neurosci 12:1062-
714 U1145.
715 Gopnik A (2020) Childhood as a solution to explore-exploit tensions. Philos T R Soc B 375.
716 Gould E, Woolley CS, Frankfurt M, Mcewen BS (1990) Gonadal-Steroids Regulate Dendritic
717 Spine Density in Hippocampal Pyramidal Cells in Adulthood. J Neurosci 10:1286-1291.
718 Humphreys KL, Lee SS, Telzer EH, Gabard-Durnam LJ, Goff B, Flannery J, Tottenham N
719 (2015) Exploration-Exploitation Strategy is Dependent on Early Experience. Dev
720 Psychobiol 57:313-321.
721 Humphries MD, Khamassi M, Gurney K (2012) Dopaminergic control of the exploration-
722 exploitation trade-off via the basal ganglia. Front Neurosci-Switz 6.
723 Johnson C, Wilbrecht L (2011) Juvenile mice show greater flexibility in multiple choice reversal
724 learning than adults. Dev Cogn Neurosci 1:540-551.
725 Johnson CM, Peckler H, Tai LH, Wilbrecht L (2016) Rule learning enhances structural plasticity
726 of long-range axons in frontal cortex. Nat Commun 7:10785.
727 Krentzel AA, Barrett LR, Meitzen J (2019) Estradiol rapidly modulates excitatory synapse
728 properties in a sex- and region-specific manner in rat nucleus accumbens core and
729 caudate-putamen. J Neurophysiol 122:1213-1225.
730 Krentzel AA, Willett JA, Johnson AG, Meitzen J (2021) Estrogen receptor alpha, G-protein
731 coupled estrogen receptor 1, and aromatase: Developmental, sex, and region-specific
732 differences across the rat caudate-putamen, nucleus accumbens core and shell. J Comp
733 Neurol 529:786-801.
734 Kwak S, Huh N, Seo JS, Lee JE, Han PL, Jung MW (2014) Role of dopamine D2 receptors in
735 optimizing choice strategy in a dynamic and uncertain environment. Front Behav
736 Neurosci 8.
737 Le Saux M, Di Paolo T (2005) Chronic estrogenic drug treatment increases preproenkephalin
738 mRNA levels in the rat striatum and nucleus accumbens. Psychoneuroendocrino 30:251-
739 260.
740 Le Saux M, Morissette M, Di Paolo T (2006) ER beta mediates the estradiol increase of D-2
741 receptors in rat striatum and nucleus accumbens. Neuropharmacology 50:451-457.
742 Lee E, Seo M, Dal Monte O, Averbeck BB (2015) Injection of a Dopamine Type 2 Receptor
743 Antagonist into the Dorsal Striatum Disrupts Choices Driven by Previous Outcomes, But
744 Not Perceptual Inference. J Neurosci 35:6298-6306.
745 Lenow JK, Constantino SM, Daw ND, Phelps EA (2017) Chronic and Acute Stress Promote
746 Overexploitation in Serial Decision Making. J Neurosci 37:5681-5689.

Lieberman OJ, McGuirt AF, Mosharov EV, Pigulevskiy I, Hobson BD, Choi S, Frier MD, Santini E, Borgkvist A, Sulzer D (2018) Dopamine Triggers the Maturation of Striatal Spiny Projection Neuron Excitability during a Critical Period. Neuron 99:540-+.

Lin WC, Delevich K, Wilbrecht L (2020) A role for adaptive developmental plasticity in learning and decision making. Curr Opin Behav Sci 36:48-54.

Matamales M, McGovern AE, Mi JD, Mazzone SB, Balleine BW, Bertran-Gonzalez J (2020) Local D2-to D1-neuron transmodulation updates goal-directed learning in the striatum. Science 367:549-+.

McCoy B, Jahfari S, Engels G, Knapen T, Theeuwes J (2019) Dopaminergic medication reduces striatal sensitivity to negative outcomes in Parkinson's disease. Brain 142:3605-3620.

Mermelstein PG, Song WJ, Tkatch T, Yan Z, Surmeier DJ (1998) Inwardly rectifying potassium (IRK) currents are correlated with IRK subunit expression in rat nucleus accumbens medium spiny neurons. J Neurosci 18:6650-6661.

Nisenbaum ES, Wilson CJ (1995) Potassium Currents Responsible for Inward and Outward Rectification in Rat Neostriatal Spiny Projection Neurons. J Neurosci 15:4449-4463.

Nonomura S, Nishizawa K, Sakai Y, Kawaguchi Y, Kato S, Uchigashima M, Watanabe M, Yamanaka K, Enomoto K, Chiken S, Sano H, Soma S, Yoshida J, Samejima K, Ogawa M, Kobayashi K, Nambu A, Isomura Y, Kimura M (2018) Monitoring and Updating of Action Selection for Goal-Directed Behavior through the Striatal Direct and Indirect Pathways. Neuron 99:1302-+.

Nussenbaum K, Hartley CA (2019) Reinforcement learning across development: What insights can we draw from a decade of research? Dev Cogn Neuros-Neth 40.

Orsini CA, Blaes SL, Hernandez CM, Betzhold SM, Perera H, Wheeler AR, Ten Eyck TW, Garman TS, Bizon JL, Setlow B (2021) Regulation of risky decision making by gonadal hormones in males and females. Neuropsychopharmacol 46:603-613.

Peak J, Chieng B, Hart G, Balleine BW (2020) Striatal direct and indirect pathway neurons differentially control the encoding and updating of goal-directed learning. Elife 9.

Proano SB, Morris HJ, Kunz LM, Dorris DM, Meitzen J (2018) Estrous cycle-induced sex differences in medium spiny neuron excitatory synaptic transmission and intrinsic excitability in adult rat nucleus accumbens core. J Neurophysiol 120:1356-1373.

Rescorla RA, Wagner AR (1972) A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Classical Conditioning II: Current Research and Theory (Black A, Prokasy W, eds), pp 64-99. New York: Appleton Century Crofts.

Sebel LE, Graves SM, Chan CS, Surmeier DJ (2017) Haloperidol Selectively Remodels Striatal Indirect Pathway Circuits. Neuropsychopharmacol 42:963-973.

Shen W, Tian X, Day M, Ulrich S, Tkatch T, Nathanson NM, Surmeier DJ (2007) Cholinergic modulation of Kir2 channels selectively elevates dendritic excitability in striatopallidal neurons. Nat Neurosci 10:1458-1466.

Sutton RS, Barto AG (1998) Reinforcement learning : an introduction. Cambridge, Mass.: MIT Press.

Tai LH, Lee AM, Benavidez N, Bonci A, Wilbrecht L (2012) Transient stimulation of distinct subpopulations of striatal neurons mimics changes in action value. Nat Neurosci 15:1281-U1152.

Tansey EM, Arbuthnott GW, Fink G, Whale D (1983) Oestradiol-17-Beta Increases the Firing Rate of Antidromically Identified Neurons of the Rat Neostriatum. Neuroendocrinology 37:106-110.

Uban KA, Rummel J, Floresco SB, Galea LAM (2012) Estradiol Modulates Effort-Based Decision Making in Female Rats. Neuropsychopharmacol 37:390-401.

Uchimura N, Cherubini E, North RA (1989) Inward Rectification in Rat Nucleus Accumbens Neurons. J Neurophysiol 62:1280-1286.

Verharen JPH, Adan RAH, Vanderschuren LJMJ (2019a) Differential contributions of striatal dopamine D1 and D2 receptors to component processes of value-based decision making. Neuropsychopharmacol 44:2195-2204.

Verharen JPH, Kentrop J, Vanderschuren LJMJ, Adan RAH (2019b) Reinforcement learning across the rat estrous cycle. Psychoneuroendocrino 100:27-31.

Wallace M, Luine V, Arellanos A, Frankfurt M (2006) Ovariectomized rats show decreased recognition memory and spine density in the hippocampus and prefrontal cortex. Brain Res 1126:176-182.

Watanabe S (2010) Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. J Mach Learn Res 11:3571-3594.

Woolley CS, Gould E, Frankfurt M, Mcewen BS (1990) Naturally-Occurring Fluctuation in Dendritic Spine Density on Adult Hippocampal Pyramidal Neurons. J Neurosci 10:4035-4039.

Xia L, Master S, Eckstein MK, Wilbrecht L, Collins A (2020) Learning under uncertainty changes during adolescence. In: CogSci.

Ye ZY, Cudmore RH, Linden DJ (2019) Estrogen-Dependent Functional Spine Dynamics in Neocortical Pyramidal Neurons of the Mouse. J Neurosci 39:4874-4888.