Estrogenic control of reward prediction errors and reinforcement learning

Carla E. M. Golden[1], Daljit Kaur[1], Andrew Mah[1], Audrey C. Martin[1], Diana H. Levy[1], Takashi Yamaguchi[2], Dayu Lin[2], Chiye Aoki[1], Christine M. Constantinople[1*]

[1] Center for Neural Science, New York University; New York, NY 10003.
[2] Neuroscience Institute, New York University Grossman School of Medicine; New York, NY 10016.

*Corresponding author. E-mail: constantinople@nyu.edu.

Abstract

Gonadal hormones act throughout the brain[1], and nearly all neuropsychiatric disorders vary in symptom severity with hormonal fluctuations over the reproductive cycle, gestation, and perimenopause[2–4]. Yet the mechanisms by which hormones influence mental and cognitive processes are unclear. Exogenous estrogenic hormones modulate dopamine signaling in the nucleus accumbens core (NAcc)[5,6], which instantiates reward prediction errors (RPEs) for reinforcement learning[7–16]. Here we show that endogenous estrogenic hormones enhance RPEs and sensitivity to previous rewards by regulating expression of dopamine reuptake proteins in the NAcc. We trained rats to perform a temporal wagering task with different reward states; rats adjusted how quickly they initiated trials across states, balancing effort against expected rewards. Dopamine release in the NAcc reflected RPEs that predicted and causally in-

fluenced subsequent initiation times. When fertile, females more quickly adjusted their initiation times to match reward states due to enhanced dopaminergic RPEs in the NAcc. Proteomics revealed reduced expression of dopamine transporters in fertile stages of the reproductive cycle. Finally, genetic suppression of midbrain estrogen receptors eliminated hormonal modulation of behavior. Estrogenic hormones therefore control the rate of reinforcement learning by regulating RPEs via dopamine reuptake, providing a mechanism by which hormones influence neural dynamics for motivation and learning.

Introduction

Estrogenic hormones, including 17β-estradiol, rhythmically fluctuate over the reproductive cycle in mammalian females and act on estrogen receptors, which can regulate protein expression via transcription and activation of signal transduction pathways[1,17,18]. These hormonal fluctuations support physiological imperatives such as homeostasis, fertility, and sexual behavior[1], but due to broad and diffuse mechanisms of action, also bind to receptors in brain regions that support cognitive behaviors, including decision-making.

One important neural system modulated by estrogen receptors is dopamine signaling in the nucleus accumbens core (NAcc), which is sensitive to application of exogenous 17β-estradiol in slices of ovariectomized animals[5,19]. The NAcc is a part of the ventral striatum that is innervated by midbrain dopaminergic neurons in the ventral tegmental area (VTA), and is required for learning cue-outcome associations[20] and determining the vigor of motivated behaviors[21]. Dopamine release in the NAcc is thought to instantiate RPEs for reinforcement learning, which is a powerful theoretical framework for describing how animals learn through interacting with

the environment[22].

In reinforcement learning, agents learn the value of taking actions in different states; those state-action values are iteratively updated based on RPEs, or the difference between received and expected rewards. A wealth of evidence suggests that dopamine release in the NAcc encodes RPEs[7–12]. Critically, manipulating dopamine release in this circuit has provided causal evidence that it acts as an RPE for reinforcement learning[13–16]. More precisely, converging synaptic inputs (e.g., from cortex, hippocampus) onto medium spiny neurons, the principal cells of the striatum, are thought to convey the state the animal is in. Dopamine is thought to mediate plasticity at these synapses, such that animals will be more (or less) likely to take actions in a state that produced positive (or negative) RPEs. In actor-critic models of the basal ganglia, the NAcc is thought to act as a "critic" that learns the expected future reward in a given state (i.e., state value)[23,24], which is known to influence response vigor[25–28]. Therefore, RPEs in the NAcc, in addition to training animals to take actions that will maximize future rewards, likely influence whether actions in a given state will be more or less vigorous[21]. The fact that estrogenic hormones can influence dopamine release in the NAcc suggests that hormones may influence RPEs for reinforcement learning. However, how behaviors unrelated to reproduction that are critical to survival, such as reward-driven learning, respond to endogenous fluctuations in estrogenic hormones is unresolved.

Here we find that rats update their response vigor on a trial-by-trial basis based on dopamine RPEs in the NAcc. We show that estrogen receptors in the midbrain enhance reinforcement learning in fertile stages of the reproductive cycle by suppressing expression of dopamine transporters in the NAcc, which increases the dynamic range of dopaminergic RPEs. This reveals a novel role for dopamine reuptake mechanisms in shaping RPEs for reinforcement learning.

Results

Rats exhibit faster learning when fertile.

We trained rats on a self-paced temporal wagering task that manipulated reward expectations by varying the magnitude of offered rewards in blocks of trials[28] (Fig. 1a,b). Rats initiated trials by poking their nose in a center port. This triggered an auditory cue, the frequency of which indicated the volume of the water reward offered on that trial (4, 8, 16, 32, or 64 μL). The reward was assigned randomly to one of two side ports, indicated by an LED. The rat could wait for an uncued and unpredictable delay to obtain the reward, or at any time could opt out by poking in the other side port to start a new trial. On 25% of trials, rewards were withheld to assess how long rats were willing to wait for them (catch trials).

To manipulate their reward expectations, rats experienced uncued blocks of trials in which they were presented with low (4, 8, or 16µL) or high (16, 32, or 64µL) rewards. These were interleaved with mixed blocks which offered all rewards (Fig. 1b). Rats initiated trials more quickly in high blocks compared to low blocks (block sensitivity, Fig. 1c). Previous work has suggested that this pattern optimally balances the costs of vigor against the benefits of harvesting reward in environments with different reward rates[29]. The sensitivity of trial initiation times to the reward blocks was robust across 306 male and female rats trained on the task (Fig. 1d).

We previously found that the decision of when to initiate a trial was governed by a reinforcement learning algorithm that updated the value of the environment or state using RPEs[28]. In actor-critic models, the critic learns the value of the state (V), or the discounted sum of future expected rewards (R) in that state, and updates state values trial-by-trial by the following equation: $V_{t+1} = V_t + \alpha(R_t - V_t)$, where $R_t - V_t$ is the RPE, and α is the learning rate. This algorithm, also called a delta rule, computes the average of previous rewards, with most recent rewards having the strongest impact. Indeed, regressing trial initiation times against previous

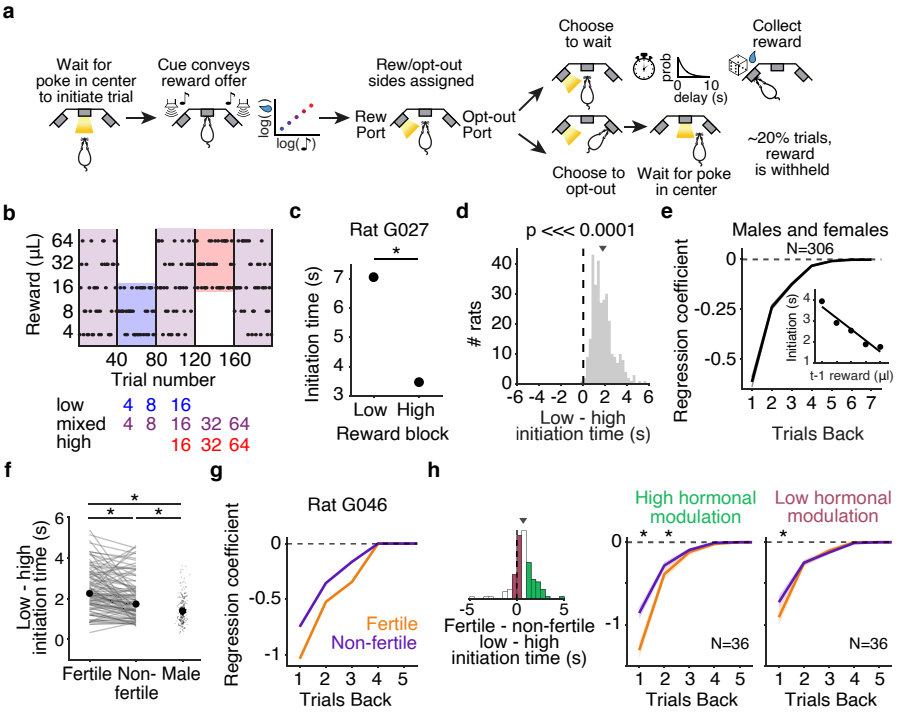

Figure 1: **Rats' trial initiation times were modulated by the value of the state and estrous stage. a.** Schematic of behavioral paradigm. **b.** Block structure of task with an example session (top) and reward distributions in each block (bottom). **c.** Mean trial initiation times for low and high blocks for one example rat are significantly different, $*p << 1 \times 10^{-20}$, two-sided Wilcoxon rank sum test. **d.** Distribution of average sensitivity of trial initiation times to blocks (low - high block) across the population is significantly different from zero, $p << 1 \times 10^{-20}$, one-tailed t-test, $N = 306$. Arrow is mean. **e.** Regression of trial initiation times against rewards on previous trials during mixed blocks for all rats. Inset is an example of initiation times plotted as a function of reward volume on one trial back for all rats. **f.** Median sensitivity of trial initiation times to blocks for 110 female rats in fertile and non-fertile stages, and 196 male rats (Kruskal–Wallis test $p = 8.13 \times 10^{-13}$ for group effect, two-sided Wilcoxon signed-rank test $p = 8.42 \times 10^{-7}$ for fertile vs. non-fertile, two-sided Wilcoxon rank sum test $p = 2.79 \times 10^{-13}$ for fertile vs. males, and two-sided Wilcoxon rank sum test $p = 7.33 \times 10^{-5}$ for non-fertile vs. males). Sensitivity $> 98^{th}$ percentile, or $< 2^{nd}$ percentile were excluded for illustration purposes. Circles are medians. **g-h.** Mean regression coefficients of trial initiation times plotted against rewards on previous trials during mixed blocks for one example rat **(g)** and for mean fertile and non-fertile stages with hormonal modulation that was high (**h** middle, fertile - non-fertile block sensitivity > 1, two-sided Wilcoxon signed-rank test, $*p = 8.09 \times 10^{-7}$ for one trial back and $*p = 0.036$ for two trials back) and low (**h** right, fertile - non-fertile block sensitivity $>$ -0.5 and $<$ 0.5, two-sided Wilcoxon signed-rank test, $*p = 1.0 \times 10^{-3}$ for one trial back). Hormonal modulation level is indicated by filled bars in the left panel of **h**. Triangle is mean.

reward offers yielded negative coefficients that declined exponentially[9], with more recent rewards having a stronger influence on behavior (Fig. 1e). When a rat received a large reward, it initiated the next trial more quickly (i.e., rewards and initiation times were negatively correlated), consistent with positive RPEs increasing state values. This suggests that rats computed an experience-based estimate of average expected rewards when deciding how quickly to initiate trials.

To determine whether hormonal fluctuations modulate reinforcement learning, we tracked female rats' reproductive (estrous) cycles with vaginal cytology (Extended Data Fig. 1a). There are four stages of the estrous cycle, each lasting $\sim$12-48 hours. 17β-estradiol is low during metestrus and diestrus and peaks during proestrus into estrus, which is when ovulation occurs (Extended Data Fig. 1b). Serum 17β-estradiol expression patterns measured with an ELISA validated stages identified by vaginal cytology (Extended Data Fig. 1c).

When female rats were in proestrus and estrus, the fertile stages of the estrous cycle, their trial initiation times were more sensitive to the reward blocks compared to when those same rats were in metestrus and diestrus, the non-fertile stages (Fig. 1f, Extended Data Fig. 1d,e). The difference between trial initiation times in low and high blocks tracked 17β-estradiol levels across the estrous cycle, with an enhanced difference in proestrus, when 17β-estradiol peaks, that continued into estrus (Extended Data Fig. 1d,e)). In non-fertile stages, when serum 17β-estradiol is lower than in fertile stages, females showed block sensitivity that was closer to males, suggesting a dose-dependent effect of estrogenic hormones (Fig. 1f, Extended Data Fig. 1b). Given that contextual effects were of a similar magnitude in proestrus and estrus, and metestrus and diestrus (Extended Data Fig. 1), we subsequently grouped these stages into fertile and non-fertile groups, respectively, consistent with previous studies[30].

In fertile stages of estrous, rats' trial initiation times were also significantly slower in both low and high blocks (Extended Data Fig. 2a). We hypothesized that this might be due to

peripheral actions of estrogen or other hormones (e.g., vasopressin), which control systemic hydration. Indeed, we measured serum osmolality in females in different stages, and found that physiological measurements of thirst varied over the estrous cycle (Extended Data Fig. 2b, c). Serum levels of 17β-estradiol were significantly negatively correlated with serum osmolality, indicating a strong correspondence between estrogenic hormones and fluid balance, consistent with previous reports of 17β-estradiol suppressing water intake and decreasing blood osmolality[31,32]. Rats exhibited lower serum osmolality, performed fewer behavioral trials, and drank less water during training sessions when they were fertile (Extended Data Fig. 2b-e). Therefore, distinct mechanisms of hormonal modulation might independently control reinforcement learning and systemic hydration.

Rats also adjusted how long they were willing to wait for the 16 μL rewards in the different blocks[28]. While there was a modest increase in block sensitivity of wait times in fertile stages, the effect size was smaller than for trial initiation times (Extended Data Fig. 3a,b). Moreover, we previously found that the wait times are governed by a different behavioral strategy that uses state inference, and that is distinct from the trial-by-trial reinforcement learning strategy that governs initiation times[28]. Because the trial initiation times reflect a reinforcement learning process that we hypothesize relies on dopaminergic RPEs, and exhibit stronger hormonal modulation, we focused on that aspect of behavior in this study.

To assess whether hormonal fluctuations modulate trial-by-trial learning at a finer timescale than reward blocks, we regressed trial initiation times in mixed blocks against previous rewards in fertile and non-fertile stages of the estrous cycle. In fertile stages, rats showed an increase in the mean previous trial coefficient (-1.066 fertile vs. -0.76 non-fertile, $p = 3.97 \times 10^{-15}$, Fig. 1g). We fit exponential curves to the coefficients for each rat, fitting to sessions from fertile and non-fertile stages separately. This revealed a stronger influence of the previous trial in fertile stages (mean value of the curve for the previous trial = -1.052 fertile vs. -0.76 non-fertile,

$p = 4.27 \times 10^{-15}$, two-sided Wilcoxon signed-rank test), and faster decay time constants (τ = 0.83 fertile vs. 0.97 non-fertile, $p = 0.021$, two-sided Wilcoxon signed-rank test), consistent with faster learning rates in standard reinforcement learning algorithms[22]. There was a broad distribution of hormonal modulation across rats (change in block sensitivity between fertile and non-fertile stages; Fig. 1h, insets). Previous trial regression coefficients from rats with high hormonal modulation (fertile - non-fertile block sensitivity > 1) showed significantly different coefficients across stages, while coefficients from rats with low hormonal modulation (absolute value of fertile - non-fertile block sensitivity < 0.5) were largely overlapping between fertile and non-fertile stages (Fig. 1h). To determine the neural mechanism, we subsequently focused on rats with hormonal modulation of block sensitivity, whose initiation times were more strongly influenced by previous rewards in fertile stages. We refrain from making claims about whether hormonal modulation of behavior, including enhanced sensitivity to previous rewards and slower overall initiation times, is better or worse for performance, since the exact benefits (costs) that are being maximized (minimized) are unobservable. Instead, we used a reinforcement learning framework to make precise descriptive statements about changes in learning over hormonal states.

Fertile stages enhance dopamine RPEs.

We next sought to characterize the neural correlates of faster reinforcement learning in rats with hormonal modulation of behavior. To test whether physiological patterns of dopamine signaling encode an RPE that is modulated by the estrous cycle, we measured dopamine in the NAcc using fiber photometry of dopamine sensors. We virally expressed the optical dopamine sensor GRAB_{DA} in the NAcc, and implanted a fiber optic over the injection site (Fig. 2a,b). We used a static fluorophore (mCherry) to correct for brain motion (Methods). We observed the strongest phasic dopamine response when the rat heard the reward offer cue, even stronger than

at the time of reward, when the rats (who were well-trained) were unambiguously cued that reward was available (Extended Data Fig. 4). Dopamine fluorescence at the offer cue scaled with the offered reward volume: there were positive phasic responses to larger rewards and negative responses to smaller rewards, consistent with an RPE (Fig. 2c,d).

We next examined the 16μL reward offers, which appeared in all reward blocks, but with a different value relative to the average reward in each block. In reinforcement learning algorithms, the value of the state (against which rewards are compared) is the discounted sum of future expected rewards in that state. In our task, the expected rewards varied with the reward blocks. The response to the auditory cue predicting 16μL was positive in low blocks, when the average reward was low, and negative in high blocks, when the average reward was high, consistent with an RPE (Fig. 2e,f). Therefore, we focused our subsequent analyses on the phasic dopamine response at the time of the offer cue. We also note that in some rats, there was an increase in dopamine as rats approached the center port, before hearing the offer cue (Fig. 2g, Extended Data Fig. 4), consistent with previous reports of approach-related dopamine release in the NAcc[16].

We found that phasic dopamine release was more strongly modulated by reward states in fertile stages, compared to non-fertile stages (Fig. 2g,h). The phasic dopamine response to the offer cue was more sensitive to the reward blocks in fertile stages of the reproductive cycle, similar to what was observed in the trial initiation time behavior (Fig. 2g). To quantify this effect, for each rat, we calculated the area under the curve of the dopamine response in the 500ms following the offer cue, for low and high blocks. For all rats, there was a larger difference in these values in fertile stages (Fig. 2h).

Dopamine release at recording sites that were proximal to the NAcc but outside of it also exhibited hallmarks of RPE encoding during the offer cue. However, dopamine at these sites was unaffected by the estrous cycle, suggesting that hormonal modulation of dopamine release

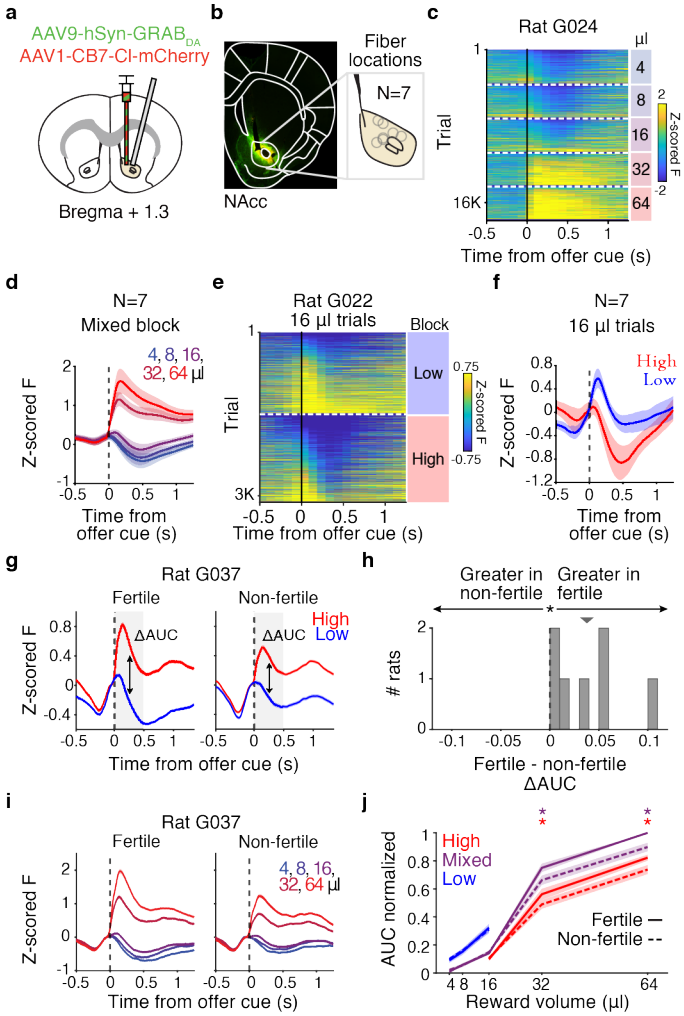

Figure 2: **NAcc dopamine at reward offer cue encodes RPEs and shows greater dynamic range in fertile stages. a.** GRAB$_{DA}$ and mCherry virus were injected into the NAcc of female rats and fibers were implanted above the injection sites. **b.** Representative example of GRAB$_{DA}$ expression in green, mCherry in red, and co-expression in yellow. Inset is fiber locations in the NAcc for 7 female rats. **c.** Heatmap of motion-corrected GRAB$_{DA}$ signal in response to reward offer cue across >16,000 trials recorded from example rat, G024, during mixed blocks and sorted by reward volume and peak response in the first 0.5 s following offer cue. Time bins are 100 ms. **d.** Average response to offer cue during mixed blocks, separated by reward volume (N=7 rats). **e.** Heatmap of motion-corrected GRAB$_{DA}$ signal in response to reward offer cue for 16 μL across >3,000 trials recorded from example rat, G022, sorted by reward block and peak response in the first 0.5 s following offer cue. Time bins are 100ms. **f.** Average response to 16 μL offer cue across the population (N=7 rats), separated by reward block.

g. Response to offer cue, separated by reward block and stage group for example rat, G037. Gray box represents window used to calculate change in area under curve (AUC), 0 to 0.5 s (ΔAUC) in **(h)**. Data is baseline-corrected using the 0.05 to 0 s before offer cue. **h.** Histogram of stage effect (fertile-non-fertile) on ΔAUC for offer cue response over rats, *$p < 0.05$, two-sided Wilcoxon signed-rank test of ΔAUC. **i.** Response to offer cue, separated by reward volume and stage group for example rat, G037, baseline-corrected using the 0.05 to 0 s before offer cue. **j.** Population average response to each reward volume, separated by reward block, min-max normalized, and baseline-corrected using the 0.05 to 0 s before offer cue, *$p < 0.05$. All error bars are mean $\pm$ S.E.M.

was specific to the NAcc (Extended Data Fig. 5).

To test whether enhanced RPEs could account for the observed behavioral effects, we used a reinforcement learning model that accounted for rats' trial initiation times[28]. In the model, initiation times were negatively correlated with state value, which was estimated with a delta rule (Fig. 3a). The model captured qualitative aspects of rats' behavior, including sensitivity of initiation times to the reward blocks and previous reward offers (Fig. 3b,c). Given that estrous had the most pronounced effects on RPEs for the larger reward volumes, 32 and 64μL, in mixed and high blocks (Fig. 2i,j), we included a multiplicative gain on large positive RPEs above the $> 90^{th}$ percentile (Fig. 3d inset). This explicitly enhanced large positive RPEs; however, because these larger RPEs increased state value, this manipulation also produced a modest increase in the dynamic range of negative RPEs (Fig. 3d), consistent with the data (Fig. 2g,i). Adding this thresholded RPE gain recapitulated the enhanced behavioral sensitivity to reward blocks in fertile stages. Moreover, it changed the impact and time constant with which previous rewards influenced the model's behavior, consistent with the change in the previous trial regression coefficients and exponential tau parameters fit to those coefficients (Fig. 1g,h).

Collectively, these data show that dopamine dynamics at the time of the offer cue encoded an RPE. Dopamine responses in the NAcc at this timepoint scaled with the offered reward (Fig. 2c,d), were inversely proportional to reward expectations (Fig. 2e,f), and exhibited a greater dynamic range in fertile stages of estrous (Fig. 2g,h). Hormonal modulation of female

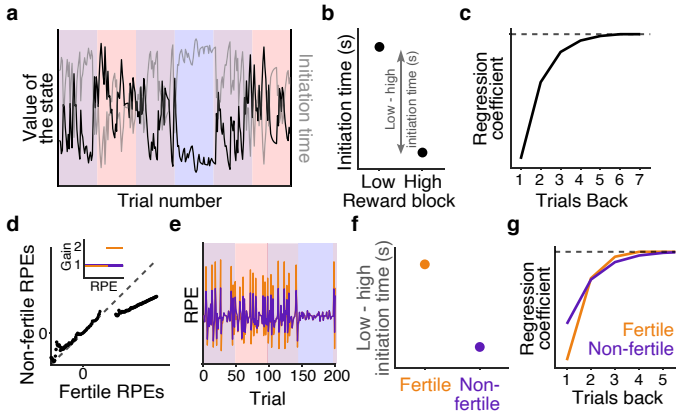

Figure 3: **Multiplicative gain on RPEs in a reinforcement learning model captures faster learning in fertile stages. a.** Simulation of the value of the state across reward blocks (blue = low, purple = mixed, red = high) using a reinforcement learning model and predicted initiation times (negatively correlated to value of the state) for one session. **b.** Model-predicted initiation times by block. **c.** Regression of model-predicted initiation times against rewards on previous trials during mixed blocks. **d.** Multiplicative gain (2x) was applied to large RPEs (RPEs $> 90^{th}$ percentile) to model enhanced RPEs observed in fertile stages. **e.** Multiplicative gain enhanced trial-by-trial RPEs, mimicking fertile stages. **f-g.** Multiplicative gain on large RPEs captures hormonal modulation of block sensitivity to reward blocks **(f)** and weight and time course of previous rewards on initiation time **(g)**.

rats' reinforcement learning could be accounted for by a multiplicative gain of RPEs, which is functionally equivalent to an increased learning rate on trials with high gain.

NAcc dopamine controls initiation times.

According to the behavioral model, dopamine RPEs update the value of the state, which determines the vigor with which animals initiate subsequent trials. To test this hypothesis, we evaluated dopamine on trials with different subsequent initiation times. In previous work, we found that initiation times were most strongly governed by RPEs following unrewarded trials[28], so we restricted the analysis of dopamine to those trials (Methods). Indeed, trials in which subsequent initiation times were fast showed larger phasic dopamine responses, and trials with long initiation times showed apparent dips in the dopamine signal (Fig. 4a). There was a

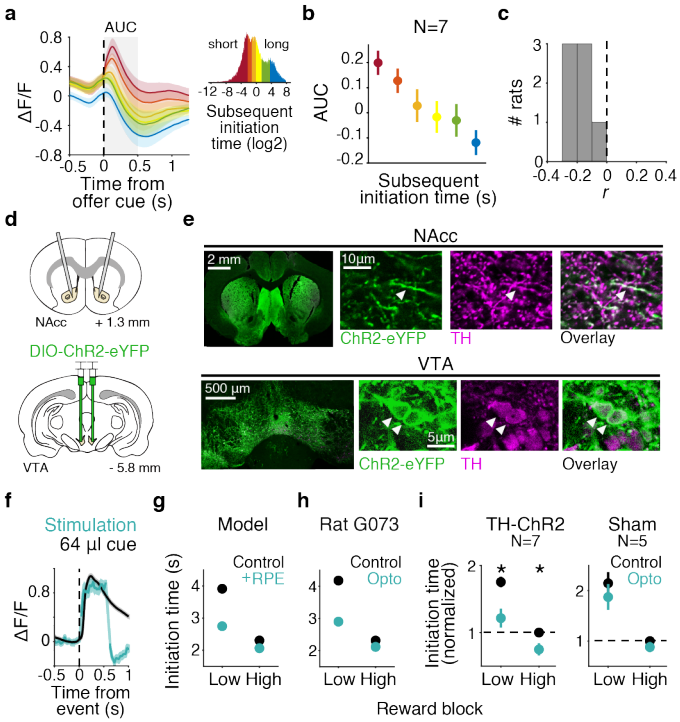

Figure 4: **Phasic dopamine release in the NAcc controls trial initiation time, consistent with an RPE that updates state values. a.** Dopamine signal aligned to the offer cue split by the subsequent trial initiation time (<0.2, 0.4, 0.9, 2.5, and 12 s and >12 s). Inset histogram shows how $\log_2$(initiation times) were binned. **b.** AUC of curves in 0.5 s window shown in **a**. **c.** Pearson correlation of AUC and subsequent initiation time over rats, all $p << 0.001$. **d.** Schematic of the strategy used to specifically stimulate TH+ VTA terminals expressing ChR2 in NAcc with 465 nm light during the offer cue on 30% of trials. **e.** ChR2-YFP colocalization with TH in VTA cell bodies and NAcc axons (20x and 63x). Arrows point to labeled NAcc axons or VTA cell bodies. **f.** Simultaneous optogenetic stimulation of TH+ axon terminals in the NAcc and measurement of $GRAB_{DA}$ fluorescence in the NAcc (blue), overlaid on top of mean response to largest reward offer cue from an example rat on mixed block trials, baseline-corrected (black). **g.** Simulated initiation times with and without an additive RPE gain on 30% of trials, using the behavioral model. **h.** Trial initiation times for an example rat for sessions with optogenetic stimulation and without (control). **i.** Mean trial initiation times across the population normalized to high blocks in control sessions for rats expressing ChR2 in TH+ axons (left) and sham controls (right, $*p < 0.05$, two-sided Wilcoxon signed-rank test). All error bars are mean $\pm$ S.E.M.

significant negative correlation between the subsequent initiation time on trial $t + 1$, and the amplitude of the dopamine on trial t, in all rats (Fig. 4b,c). In other words, the phasic dopamine response at the offer cue predicted the rats' initiation time on the next trial, consistent with dopamine at this event acting as an RPE to update the value of the state for response vigor.

To test whether dopamine at this timepoint causally influenced initiation times, we used optogenetics to activate dopamine terminals in the NAcc. We expressed Cre-dependent excitatory opsins (AAV9-DIO-ChR2.0) bilaterally in the VTA and substantia nigra of tyrosine hydroxylase (TH)-Cre rats and rats injected with a TH-Cre virus, and bilaterally implanted fiber optics over the NAcc (Fig. 4d,e). We histologically verified that ChR2 expression was largely restricted to neurons that were TH-immunoreactive (Fig. 4e, Extended Data Fig. 6). On 30% of trials, we optogenetically stimulated midbrain dopamine terminals in the NAcc triggered when rats heard the tone indicating the reward offer. Simultaneous stimulation and photometry measurements of GRAB_{DA} indicated that optogenetic stimulation elicited dopamine release comparable to evoked transients during large reward offers (Fig. 4f).

We reasoned that this manipulation was similar to providing an additive gain to RPEs since it would add an (approximately) constant amount of dopamine on stimulation trials. Simulating an additive RPE gain in the behavioral model predicted that rats should exhibit faster trial initiation times because positive prediction errors would increase state values (Fig. 4g). The model predicted a larger effect in low blocks, because it implemented a floor effect to capture the fact that rats can only initiate trials so quickly in high blocks. Indeed, we found that dopamine terminal stimulation decreased trial initiation times (Fig. 4h,i). Notably, there was no effect on trial initiation times in sham animals that did not express light-sensitive opsins (Fig. 4i). There was also no effect on wait times (two-way ANOVA tests, $F = 0.24$ and $p = 0.91$ for an interaction between reward and session type for TH-Cre rats and $F = 0.207$ and $p = 0.103$ for sham rats), further showing that dopamine at reward offer cue encodes an RPE that is selectively

used to guide initiation times. Collectively, these results support the hypothesis that dopamine signaling in the NAcc at the time of the reward offer acts as an RPE to adjust the value of the state for subsequent trial initiation times.

Reduced dopamine reuptake in fertile stages.

To identify the cellular mechanisms by which dopamine signaling was enhanced in fertile stages of the estrous cycle, we performed quantitative mass spectrometry of the NAcc in diestrus, proestrus, and estrus (N=5 rats in each group). 238 proteins were significantly decreased in expression in fertile stages compared to non-fertile (diestrus), and 258 increased. Gene ontology analysis identified several proteins involved in dopamine signaling that were differentially expressed in fertile and non-fertile stages (Methods, Extended Data Fig. 7). However, the only proteins that were differentially expressed in a way that could account for the observed increase in phasic dopamine in fertile stages were proteins involved in dopamine reuptake, which were expressed at lower levels in these stages (Fig. 5a, Extended Data Fig. 7). The dopamine transporter (DAT or Slc6a3) controls dopamine reuptake in the NAcc, and was expressed at significantly lower levels in proestrus compared to diestrus. The serotonin transporter (SERT or Slc6a4) is also capable of robust dopamine reuptake[33], and was expressed at lower levels in proestrus and estrus compared to diestrus. Smpd3 (sphingomyeline phosphodiesterase 3) triggers fusion of DAT to the plasma membrane, and has been shown to make dopamine reuptake more efficacious[34,35]. Smpd3 was decreased in expression in estrus, suggesting a lower density of DAT in the plasma membrane during estrus, consistent with effects of 17β-estradiol applications on DAT localization in vitro[35]. Reduced plasmalemmal DAT should in principle lead to higher levels of extracellular dopamine.

We verified that rats exhibited reduced expression levels of DAT and SERT in fertile stages of the estrous cycle, using a different assay, immunohistochemistry, and a separate cohort of

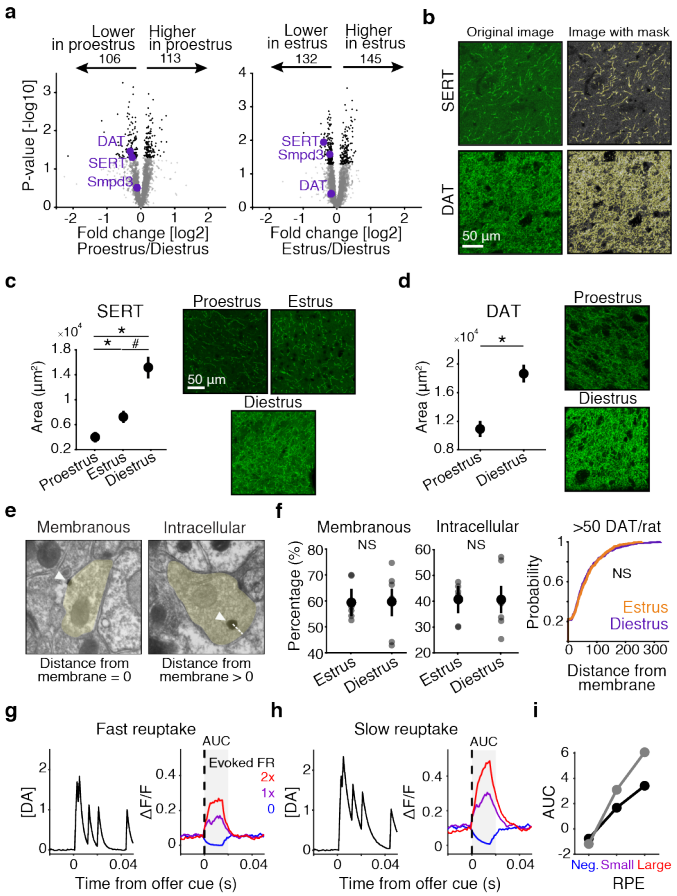

Figure 5: **Dopamine reuptake proteins are decreased in expression in fertile stages. a.** Volcano plot depicting significantly differentially expressed proteins (with FDR correction of 5%) in black. Proteins related to dopamine reuptake that were decreased in expression in fertile stages are highlighted in purple. **b.** Example images from immunohistochemical quantification of SERT and DAT expression at 63x with mask identifying fluorescent pixels above a threshold, generated in ImageJ, overlaid. **c-d.** Area of mask segmentation for SERT (**c**, $*p < 0.01$ Kruskal-Wallis test for group effect and $*p < 0.05$ and $\#p < 0.1$ Wilcoxon rank sum post-hoc tests) and DAT (**d**, $*p << 0.01$ two-sided Wilcoxon rank sum test) in fertile and non-fertile stages, with example images from each stage. **e.** Example of DAT labeled with silver-intensified gold and imaged with electron microscopy. Arrows point to putative DAT proteins. Minimum distance between plasma membrane and center of DAT particle are shown with dotted lines. Axons are highlighted in yellow. **f.** Proportion of membranous and intracellular DAT particles quantified from electron microscopy images. Third panel is distance from the membrane, expressed as a cumulative probability. **g-h.** Simulation of dopamine concentrations following action potentials with firing rates (FR) that scale with RPE (left panels) with fast **(g)** or slow **(h)** reuptake and dopamine fluorescence for negative, small, and large RPEs (right panels).

i. Quantification of AUC in **g-h**. All error bars are mean ± S.E.M.

rats in diestrus, proestrus, and estrus (N=3 per group). All sections were co-incubated in the same solutions simultaneously, to ensure that differences in fluorescence did not reflect subtle differences in tissue processing. We quantified fluorescently labeled pixel area from confocal photomicrographs, where fluorescence indicated DAT or SERT immunoreactivity (Fig. 5b). Consistent with the proteomics results of reduced DAT and SERT expression in fertile stages, we found that significantly fewer pixels were DAT- and SERT-immunoreactive in proestrus and estrus compared to diestrus (Fig. 5c-d). It is notable that these results held across different assays (proteomics and immunohistochemistry) and cohorts of rats.

Smpd3 is implicated in triggering fusion of DAT to the plasma membrane, and was also decreased in expression in estrus. To test whether there was also a change in subcellular localization of DAT in estrus, we characterized DAT localization at the ultrastructural level. Individual DAT particles were labeled with silver-intensified gold and visualized with electron microscopy. We quantified the fraction of DAT particles that were continuous with the membrane (membranous) versus discontinuous (intracellular), and also measured the distance from the center of the intensified gold particle to the plasma membrane. There was no significant difference in the fraction of DAT particles that were membranous versus intracellular, nor was there a difference in DAT particles' proximity to the membrane between estrus and diestrus (Fig. 5e,f). Therefore, reduced levels of DAT and SERT expression, but not differences in subcellular localization of DAT, likely accounted for enhanced dopamine RPEs and learning in fertile stages.

How could changes in reuptake, which ostensibly would result in elevated levels of extracellular dopamine, account for the enhanced phasic responses we observed? We created a computational model in which action potentials of dopamine neurons caused instantaneous increases in dopamine concentration, which decayed with a rate that reflected reuptake dynamics, unless another action potential caused another instantaneous increase. The fluorescent signal at

each time step was treated as a noisy observation of the underlying dopamine concentration[36]. The model assumed that the Poisson spike rate of the dopamine neurons increased or decreased (to zero) for positive and negative RPEs, respectively. When the dopamine concentration decayed more slowly, to simulate reduced reuptake, this increased the likelihood that dopamine from consecutive action potentials would temporally summate before decaying back to baseline. Averaged over many trials, this temporal summation mechanism was sufficient to produce larger positive phasic responses, consistent with what we observed by measuring dopamine release in fertile stages (Fig. 5g,h). Moreover, increasing temporal summation modestly elevated baseline levels of dopamine, producing a greater dynamic range for negative RPEs, which is normally thought to be limited by the low background firing rates of dopamine neurons[9] (Fig. 5g,h). These results are consistent with previous studies showing that inhibition of DAT increases both tonic and phasic dopamine[37,38]. In our simulation, reduced reuptake produced the strongest enhancement of the large phasic DA responses, similar to our empirical observation that RPEs for large rewards were most strongly enhanced in fertile stages. Therefore, decreased expression of DAT could produce a greater dynamic range of phasic dopaminergic RPEs by promoting temporal summation of dopamine transients from consecutive action potentials.

Reduced dopamine reuptake would presumably result in elevated levels of baseline or "tonic" dopamine levels, which has been hypothesized to determine animals' overall response vigor[25,39]. However, we found that in fertile stages of estrous, when DAT and SERT were expressed at lower levels, rats' trial initiation times were significantly slower in both low and high blocks (Extended Data Fig. 2), suggesting reduced response vigor. Our findings of reduced blood osmolality suggest that this might be due to peripheral actions of estrogen or other hormones (e.g., vasopressin), which control systemic hydration. Therefore, reduced dopamine reuptake is not sufficient to increase overall response vigor, but rather likely interacts with other circuits and physiological systems to determine motivation and vigor, as has been shown previously with

food-deprivation status[40].

ERα controls hormonal modulation of behavior.

17β-estradiol is one of many hormones whose levels fluctuate over the estrous cycle. We next sought to test whether estrogenic signaling was necessary for hormonal modulation of reinforcement learning over the reproductive cycle. Expression of DAT in the NAcc is restricted to dopaminergic axons from the midbrain. Therefore, we hypothesized that estrogen receptors, which can regulate protein expression via transcription and other mechanisms when bound to 17β-estradiol[1,17,18], might influence expression of DAT in dopaminergic neurons. We virally expressed short-hairpin RNA (shRNA), labeled with eYFP, to knockdown expression of the estrogen receptor ERα bilaterally in the VTA (shEsr1-eYFP, Fig. 6a). Viral expression was inducible by doxycycline in the diet and water, allowing for flexible control of estrogen receptor expression in the midbrain. Post-mortem histology confirmed reduced expression of ERα at the site of virus expression in rats consuming doxycycline (Fig. 6b,c). Genetic suppression of estrogen receptors in the midbrain strongly reduced hormonal modulation of behavior in fertile stages, when there are high levels of circulating estrogenic hormones (Fig. 6d,e). In rats that showed enhanced sensitivity to reward blocks in fertile stages, suppression of estrogen receptors reduced block sensitivity to that of non-fertile stages. The same regimen of doxycycline had no effect on control rats not expressing the inducible shEsr1 virus (Kruskal–Wallis test for group effect $p = 0.643$.). Therefore, we conclude that midbrain estrogen receptors mediate hormonal modulation of reinforcement learning (Fig. 6f).

Discussion

Using a task with a continuous behavioral read-out of trial-by-trial reinforcement learning, we found that female rats learn more from previous rewards when they are in fertile stages of

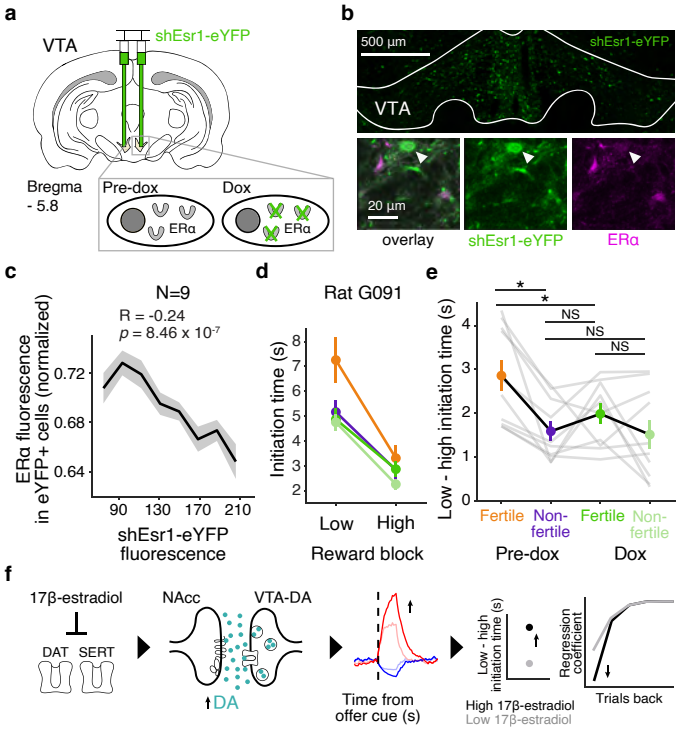

Figure 6: **Midbrain estrogen receptor (ERα) knockdown suppresses hormonal modulation of behavior. a.** shEsr1-eYFP was injected bilaterally in the VTA and expression was induced with doxycycline (dox). **b.** Expression of eYFP-labeled shEsr1 virus in the VTA (20x, top). Reduction in expression of ERα in eYFP-labeled cells compared to cells not expressing eYFP (63x, bottom). Arrow points to lack of ERα expression in eYFP-labeled cell. **c.** Mean ERα fluorescence in eYFP-labeled cells, normalized to mean ERα signal, plotted as a function of shEsr1-eYFP viral expression across photomicrographs from 9 rats. Statistics are from a Pearson correlation. **d.** Mean and S.E.M. trial initiation time in low and high blocks in fertile and non-fertile stages before induction of the virus (pre-dox) and following induction with dox (during dox) for an example rat. **e.** Average sensitivity of trial initiation times to reward block for 10 female rats before and after dox (Kruskal–Wallis test $p = 0.037$ for group effect, two-sided Wilcoxon signed-rank tests for post-hoc tests: $*p = 0.002$ for pre-dox fertile vs. pre-dox non-fertile, $*p = 0.037$ for pre-dox fertile vs. dox fertile, $p = 0.23$ for pre-dox non-fertile vs. dox fertile, $p = 0.77$ for pre-dox non-fertile vs. dox non-fertile, and $p = 0.56$ for dox fertile vs. dox non-fertile). **f.** Schematic summarizing findings. 17β-estradiol suppresses expression of transporters that reuptake dopamine. This increases the amount of dopamine in the extracellular space, allowing for a greater dynamic range of RPEs and enhanced learning from previous reward in fertile stages. All error bars are mean $\pm$ S.E.M.

the estrous cycle and compared to males, building on previous studies that found sex-dependent effects of reward on motivation[30]. We reveal a computational and circuit-level mechanism that underlies estrous-dependent enhanced learning for which estrogenic hormones are likely the master regulator: expression of the reuptake proteins DAT and SERT in the NAcc is reduced in fertile stages of the estrous cycle, resulting in a larger dynamic range of dopamine levels that enhances RPEs, which causally influence initiation times on subsequent trials (Fig. 6f). Our findings suggest that the enhancement is most pronounced for large positive RPEs, possibly due to increased occupancy of pools of DAT and SERT proteins at high dopamine concentrations. Surface-level expression of DAT and SERT in the striatum is highly dynamic[41,42], with protein turnover occurring on the order of hours to days[42–44], suggesting that hormonal fluctuations over estrous could plausibly affect plasmalemmal transporter expression.

One key future question is whether fertile stages also modulate the firing rates of dopaminergic cells in the VTA, in addition to dopamine dynamics in the NAcc. While previous studies have examined estrous-dependent VTA dynamics in the context of drug exposure[45,46], it is unclear whether estrogenic hormones affect encoding of RPEs during reinforcement learning at the level of spikes in the VTA. Estrogenic hormones could in principle act directly on VTA neurons[47,48], or have indirect effects by modulating the excitability of afferent inputs to the VTA[49]. Indeed, medium spiny neurons in the NAcc are known to project back to the VTA[50], so it is possible that the estrogenic modulation of dopamine dynamics in the NAcc described here could indirectly influence VTA activity, via this projection. Distinct subpopulations of dopamine neurons in the VTA broadcast RPEs to different downstream circuits[51], so a question for future exploration is whether estrous-dependent modulation of RPEs is circuit-specific, and largely restricted to state value learning in the NAcc, or includes the dorsal striatum. We found that dopamine release at sites that were adjacent to the NAcc, including the NAc shell and ventromedial striatum, were not modulated by estrous, suggesting regional specificity of

estrogenic modulation. Local modulation of reuptake may provide more spatially precise control of dopamine signaling compared to regulation at the level of somatically-generated action potentials, given the enormous axonal arborizations of dopamine neurons[52].

Here, we find a novel role for dopamine transporters in shaping RPEs for reinforcement learning. Drugs of abuse such as cocaine and methamphetamine act on DAT[53], and both DAT and SERT are the targets of pharmacological treatments for many neuropsychiatric disorders including schizophrenia, depression, anxiety, and attention deficit hyperactivity disorder[54]. Computational psychiatry models have examined these disorders through the lens of reinforcement learning[55–57]. Our results provide a potential biological mechanism that bridges dopamine reuptake and reinforcement learning in health and disease, and might also underlie hormonal modulation of symptom severity in females[2–4].

There is a major gap in our understanding of how endogenous hormones interact with neuromodulatory systems on fast timescales during rich, quantifiable behaviors. We developed a reinforcement learning model that accounted for rats' behavior, and that included a computation that dopamine is thought to instantiate in the brain. Our computational and decision-theoretic approach provided a principled framework for interpreting behavioral, physiological, optogenetic, and molecular profiling data, and represents a novel approach for studying endocrine systems in the brain. This approach allowed us to characterize estrogenic modulation of dopamine at broad levels of description, and opens up new avenues of research of hormone-neuromodulatory interactions during complex cognitive behaviors.

Methods

Subjects A total of 362 Long-evans rats (201 male, 161 female) between the ages of 6 and 24 months were used for this study (*Rattus norvegicus*), including TH-Cre (N =23), ADORA2A-Cre (N =10), and DRD1-Cre (N =3). Animal use procedures were approved by

the New York University Animal Welfare Committee (UAWC #2021-1120) and carried out in accordance with National Institutes of Health standards.

Rats were pair-housed when possible, and water restricted to motivate them to perform behavioral trials. From Monday to Friday, they obtained water during behavioral training sessions, which were typically 90 minutes per day, and a subsequent ad libitum period of 20 minutes. Following training on Friday until mid-day Sunday, they received ad libitum water. Rats were weighed daily.

Behavioral training

We have previously published a detailed description of behavioral training for this task[28]. Reward blocks were 40 completed trials, in which the rat remained nose poking in the center port for a variable period randomly drawn from a uniform distribution from [0.8, 1.2] seconds. On each trial, an auditory tone [1, 2, 4, 8, 16kHz] signaled one of the five different rewards. There was a one-to-one mapping between tones and reward volumes, from smallest to largest. On each trial, the reward delay period was randomly drawn from an exponential distribution with a mean of 2.5 s; the catch probability was 15-35% until the rat matched the catch rate with opt-out rate, and then held at 25%.

Early cohorts of female rats experienced the same reward set as the males [5, 10, 20, 40, 80μL]. However, female rats consumed less water due to their smaller body size and performed substantially fewer trials than the males. Reward offers for female rats were slightly reduced while maintaining the logarithmic spacing to [4, 8, 16, 32, 64 μL] to obtain a sufficient number of behavioral trials. For analysis, reward volumes were treated as equivalent to the corresponding volume for the male rats.

Criteria for including behavioral data

To determine when rats were sufficiently trained to understand the mapping between the auditory cues and water rewards, we evaluated their wait times on catch trials as a function of offered rewards. Only sessions with a running average slope (using a window of 5 sessions) above 0.01 (the average slope of shuffled data) were included, to accommodate cycle-dependent changes in task performance. We emphasize that the criteria for including sessions in analysis did not evaluate rats' sensitivity to the reward blocks, or their trial initiation times.

Estrous cycle tracking

We tracked the estrous cycle for all females included in this study using vaginal cytology, with vaginal swabs collected immediately after each session using a cotton-tipped applicator first dipped in de-ionized water. Samples were smeared onto a clean glass slide and visually classified under bright-field illumination on a Leica ICC50 E microscope at 10X. The estrous cycle stage on each day was manually determined on the basis of the proportion of leukocytes, cornified epithelial cells, and nucleated epithelial cells according to established protocols using unstained smears[58]. Given that experiments were performed over the course of months, days to weeks of irregular cycling were captured and included in the data presented. Once the rat entered reproductive senescence, determined by a continuous estrus phase lasting longer than two weeks, all sessions during that time were excluded from the analyses.

To validate the accuracy of our visual classification, we measured 17β-estradiol in serum from trunk blood, collected in serum-separating tubes and spun at 3,000 rpm at 4°C for 10 min, with the mouse/rat estradiol ELISA kit from Calbiotech (ES180S-100) after determining stage with the method described above in 18 rats (4 in proestrus $>$ 5 hours before lights out; 4 in proestrus within 3 hours before lights out; 5 in estrus; 5 in diestrus), run in triplicates.

Behavioral model

We modeled the trial initiation times as being negatively correlated to state value according to the following equation[28,29]:

$$init_t = -mV_t + b, \tag{1}$$

where V_t represents the recency-weighted average of all previous rewards. This average depends on the learning rate parameter α with the recursive equation

$$V_{t+1} = V_t + \alpha\delta_t, \tag{2}$$

where V_t is the value of the state on trial t, r_t is the reward on trial t, $\delta_t = r_t - V_t$ is the RPE, and α is a learning rate between 0 and 1.

Dopamine reuptake model

We modeled the activity of dopamine neurons as Poisson spikes with a background firing rate that increased or decreased (to zero) to reflect positive or negative RPEs, respectively. The dopamine concentration exhibited instantaneous increases at action potentials, which decayed over time to reflect reuptake, according to the following equation:

$$c_t = \gamma c_{t-1} + z_t, \tag{3}$$

where c_t is the dopamine concentration, γ determines the rate of decay or reuptake, and $z_t = 1$ indicates the presence of a spike at time t, or 0 otherwise. The measured dopamine fluorescence at each timepoint, y_t, was treated as an observation of the true concentration with Gaussian noise:

$$y_t = c_t + \epsilon. \tag{4}$$

We simulated 500 trials for each condition (positive and negative RPEs), and averaged the measured dopamine signal over those trials. We performed this simulation for $\gamma = 0.65$, corresponding to faster reuptake, and $\gamma = 0.85$, corresponding to slower reuptake.

Histological verification and protein staining

Rats were euthanized following completion of behavioral testing for photometry, optogenetics, and shRNA experiments and when they reached adulthood (10 - 11 weeks) for electron microscopy and immunohistochemistry quantification experiments. Rats were first perfused with phosphate buffer saline (PBS) and subsequently with 4% paraformaldehyde for DAT and SERT quantification, and formalin for histological verification for optogenetics, photometry, and shEsr1 experiments. Brains were post-fixed with 4% paraformaldehyde for one week for DAT and SERT quantification and with formalin for two days for all other experiments. 50 μm sections of extracted brain tissue were sliced on a Leica VT1000 vibratome.

Sections were treated with 1% hydrogen peroxide in 0.01 M PBS for 30 min, washed with 0.01 M PBS, and blocked in 0.01 M PBS containing 0.05% sodium azide and 1% bovine serum albumin (PBS-BSA-Azide) for 30 minutes before being treated with primary antibody overnight at room temperature (RT). Primary antibody was removed with PBS rinses and then sections were treated with secondary antibodies for 1 hour away from light. Primary and secondary antibodies were made in PBS-BSA-Azide and primary antibodies additionally contained 0.2% triton-x. To verify virus expression and implant location, endogenous fluorescent markers were amplified with rabbit anti-GFP (Thermo Fisher Scientific A11122, 1:2,000, lot #2083201) and imaged with an Olympus VS120 Virtual Slide Microscope. For light microscopy experiments, DAT was labeled with 1:500 Sigma-Aldrich D6944 lot #0000124874, SERT was labeled with 1:2,000 Chemicon AB1594P lot #18020403, TH was labeled with 1:400 Millipore MAB318 lot #3990619, and ERα was labeled with 1:250 Thermo Fisher Scientific PA1-309 lot #YG378288.

To increase the immunoreactivity of ERα in the VTA after fixation, antigen retrieval was conducted with citrate buffer. The secondary antibodies used were as follows (all 1:200 and from Thermo Fisher Scientific): anti-rabbit AlexaFluor 488 IgG A11008, lot #2147635) for DAT, SERT, and GFP, goat anti-mouse AlexaFluor 555 for TH (A21424, lot #2390715), and goat anti-rabbit AlexaFluor 594 (A11037, lot #2160431) for ERα. All slices were mounted on glass slides using Prolong Gold (Vector Laboratories).

Quantification of fluorescence

Photomicrographs were taken at 20x and 63x with a Leica DM6 CS Confocal using standardized parameters developed to maximize signal-to-noise ratios. A mask was generated to identify the location of fluorescing pixels. The total area of the mask (μm^2) was used to determine expression level of DAT and SERT. Mean fluorescence of ERα-labeled pixels that were also labeled with GFP were normalized to mean fluorescence of all ERα-labeled pixels. Normalized fluorescence for ERα and mean fluorescence of GFP were used in the analysis.

Stereotaxic surgeries

All surgeries were performed in female rats after 4 months of age with a Neurostar Robot Stereotaxic, using 3% isofluorane in oxygen at a flow rate of 2.5 L/minute for induction and 2% isofluorane in oxygen at a flow rate of 1.75 L/minute as maintenance for the duration of the procedure. VTA injections were conducted at a 10° angle from the midline and targeted AP -5.8; ML 0.4-0.6; DV -7.8-8.3. NAcc injections and implants targeted AP 1.3; ML 1.65; DV -6.9 with an 8-10° angle from the midline for bilateral implants.

Photometry

We used fiber photometry and the $GRAB_{DA}$ dopamine sensor (AddGene #140554) to measure dopamine activity during the task. AAV9-hsyn-GRAB_DA2h was injected into the NAcc. To

control for motion artifacts, mCherry (AAV1-CB7-CI-mCherry-WPRE-RBG, AddGene #105544) was also expressed because it is constitutively active. 60 nl of equal parts $GRAB_{DA}$ and mCherry were delivered in each injection. Chronically implantable optic fibers (Thor labs) with 400 μm core, 0.5 NA fiber optics were implanted unilaterally over the injection site (DV -6.7). Doric Lenses hardware and software (Doric Neuroscience Studio) were used to record fluorescence. Two-channel motion artifact correction (TMAC) was used to correct for movement artifacts, with mCherry as the activity-independent channel[59].

Optogenetics

We photoactivated VTA axons in NAcc by bilaterally expressing AAV9-EF1a-double floxed-hChR2 (AddGene #20298), 500 nl per injection, in VTA of TH-Cre rats (N=5) or simultaneously expressing AAV9-rTH-PI-Cre (AddGene #107788, N=3, 250 nl ChR2 and 250 nl TH-Cre) to restrict expression to dopaminergic projections. Two tapered fibers (0.66 NA, 1 mm emitting length, 11 mm implant length, 200 μm core diameter, Optogenix) were implanted in stimulation and sham rats as controls (N=5). Continuous blue light (465 nm PlexBright LEDs, Plexon) was delivered randomly on 30% of trials for 500-700 ms during the center poke, reaching 7.5 mA.

To measure the effect of photoactivation of VTA DA axons in the NAcc on local dopamine release, dopamine was recorded with a red fluorescent $GRAB_{DA}$ sensor (AddGene #140557-AAV9-hSyn-GRAB_rDA1h), while VTA DA terminals in the NAcc were stimulated for the approximate duration of the offer cue (500-600 ms). The relative increase in dopamine was compared to the mean change in dopamine during the offer cue for all trials with the largest reward for an example rat (G051, 64 μl cue) .

Relative protein abundance profiling

16 rats (5 in diestrus, 5 in proestrus, and 6 in estrus), anesthetized with isofluorane, were decapitated and a brain matrix (Braintree Scientific) was used to make 1.0 mm coronal brain slices containing the central aspects of the NAcc (0–1.6 anterior to Bregma), kept on wet and dry ice. Bilateral 15 gauge punches of the NAcc were obtained and immediately placed on dry ice.

Samples were reduced with dithiothreitol at 57° C for 1 hour (5 μl of 0.2 M), alkylated with iodoacetamide at RT in the dark for 45 minutes (5 μl of 0.5 M). Urea concentration was lowered to ¡2 M by addition of 100 mM Ammonium Bicarbonate and sequencing-grade modified trypsin 1.5 μg (Promega) was added to each sample. Digestion proceeded overnight on a shaker at RT. To extract proteins, samples were de-salted with Sep-Pak C18 Cartridges. C18 Cartridges were conditioned by passing 1 ml of 90% acetonitrile in 0.5% acetic acid solution three times, then they were equilibrated by passing 3 ml of 0.1% TFA solution through them two times. Acidified samples were slowly loaded onto corresponding C18 cartridge packing. Samples were de-salted by passing 1 ml of 0.1% TFA solution through the cartridge three times. Then the cartridge was washed with 1 ml of 0.5% acetic acid. Peptides were eluted into clean tubes by slowly passing 500 μl of 40% acetonitrile in 0.5% acetic acid, followed by the addition of 80% acetonitrile in 0.5% acetic acid.

Samples were reconstituted in 30 μl of 100 mM HEPES solution, pH 8.5 and labeled with TMT dissolved in 90 μl of 100% anhydrous EtOH for 5 min. Each sample was combined with its corresponding TMT label and incubated for 1 hour. The reaction was quenched by adding 50 μl of 100 mM ammonium bicarbonate solution and incubating for 15 min. Samples were dried in a SpeedVac concentrator. The sample was then fractionated into 45 fractions using basic HPLC fractionation and the fractions were concatenated into 15 final fractions for LC-MS analysis where $1/10^{th}$ of the fractions were analyzed individually. LC separation online with MS using the autosampler of an EASY-nLC 1000 (Thermo Scientific). Peptides were gradient eluted from

the column directly to Eclipse mass spectrometer using a 1 hour gradient (Thermo Scientific) using 2% acetonitrile and 0.5% acetic acid for solvent A and 80% acetonitrile and 0.5% acetic acid for solvent B.

High resolution full MS spectra were acquired with a resolution of 60,000, an AGC target of 4e5, with a maximum ion time of 50 ms, and scan range of 400 to 1500 m/z. Following each full MS top ten data dependent HCD MS/MS spectra were acquired. All MS/MS spectra were collected using the following instrument parameters: resolution of 60,000, AGC target of 1e5, maximum ion time of 60 ms, one microscan, 0.7 m/z isolation window, and NCE of 30. MS/MS spectra were searched against the rat database using MaxQuant.

Only proteins that had at least two peptides detected and that were identified in at least three samples of one estrous stage were included in the analysis. 3082 proteins were identified and 3043 were quantified. Reporter ion intensity values were normalized and all protein intensities were Log2 transformed. Proteins were determined to be significantly differentially expressed with a t-test and a Benjamini-Hochberg false discovery rate (FDR) of 5% to correct for multiple comparisons.

Electron microscopy

12 rats were anesthetized with isofluorane and perfused through the ascending aorta with 50 mL of PBS containing 0.2 mL of heparin and then 4% paraformaldehyde dissolved in 0.1M phosphate buffer, pH 7.4. The brains were removed from the calvariae and post-fixed for one week in 4% paraformaldehyde. Sections through the NAcc (50 μm) were cut with a vibratome, and stored in 0.01 M PBS and 0.05% azide. To enhance the antibody's penetration of tissue, free-floating vibratome sections underwent a freeze/thaw protocol, for which tissue was cryoprotected in increasing concentrations of dimethyl sulfoxide (DMSO; 5%, 10%, and 20%) for 10 minutes each before rapid freezing in a beaker of 4-methylbutane chilled using dry ice and

100% ethanol (EtOH) as the heat conduit. After rapid freezing, the tissue was thawed by immersing in a bath of 20% DMSO at RT, before another bout of rapid freezing for a total of 8 freeze/thaw sequences. Tissue was then incubated in decreasing concentrations of DMSO (10% and 5%) for 5 minutes each before returning to PBS[60]. Tissue was stained for DAT using the protocol described above, except with an anti-DAT concentration of 1:250 that the tissue was treated with for 48 hours. Secondary antibody incubation was overnight in PBS-BSA-Azide containing ultra-small colloidal gold-conjugated goat anti-rabbit secondary antibody (1:100, EMSciences #25181, lot #GAR-01217/1). Sections were rinsed in PBS and post-fixed in 2% glutaraldehyde in PBS for 10 min, and then the steps for silver intensification (KPL Silver Enhancer Kit, #5520-0021) were carried out to enlarge the ultra-small colloidal gold particles.

The EM tissue processing was similar to the procedure performed before[61], which consisted of osmium-free tissue processing in 1% tannic acid, 1% uranyl acetate and 0.2% iridium tetrabromide, each dissolved in maleate buffer (pH 6.0, 0.1 M), post-fixation in 1% uranyl acetate in 70% EtOH overnight, dehydration in 70%, 90%, and 100% EtOH, followed by 100% acetone, infiltration in EPON 812 (EM Sciences), flat-embed between Aclar plastic sheets (EMSciences) and capsule-embedded in BEEM capsules (EMSciences). Ultrathin sections were prepared using the Ultracut E ultratomicrotome and collected on nickel grids (EMSciences). Sections were counter-stained with lead citrate to improve contrast. Images were captured at a magnification of 50,000× using the Hamamatsu CCD camera (1.2 megapixel) and software from AMT (Boston, MA, USA).

For each animal, approximately 50 axons were sampled and quantified. Axons were considered to be DAT-positive when one or more silver-intensified gold (SIG) particles could be detected along its plasma membrane or within its cytosol. Axons were identified to be forming symmetrical or asymmetrical synapses. SIG particles were categorized as either membranous (center of particle was perfectly aligned to plasma membrane) or intracellular (center of particle was

within the boundaries of the axon). The size of the axons and SIG particles were assessed by measuring the maximum and minimum diameters. The distance of each SIG particle to the plasma membrane was measured using the center of the SIG particle and the nearest point along the membrane. 897 SIG particles were included in the analysis.

Estrogen receptor suppression

AAV2-V6/TO-shEsr1-v2-CMV-TeT-eGFP (shEsr1) was injected bilaterally into the VTA in 10 rats to knockdown ERα expression[62]. Doxycycline (dox) was used to induce viral expression of shEsr1, delivered via both food and water. Dox-containing food (Bio-Serv diet) was available ad lib and dox-containing water (3 mg/mL Sigma-Aldrich dox powder) was available for 10 minutes, totalling a daily dosage of 20 mg/kg (below toxic levels)[63]. Dox-containing water was followed by 10 minutes of regular water ad lib to maintain comparable thirst levels. All behavioral sessions after the first day of the dox regimen were included in the analysis and compared to sessions that preceded surgery. Rats that received the dox regimen but did not receive the shEsr1 injections were used as controls.

Serum osmolality

Osmolality was measured in the same serum samples used to detect 17β-estradiol expression. The extracted serum was sent by refrigerated transportation to a commercial laboratory for osmolality determination using a freezing-point method (Cornell University Clinical Pathology Laboratory). Osmolality measurements $> 2.5\times$ the standard deviation above the mean were considered outliers and excluded. 17β-estradiol expression for one rat in estrus was excluded when correlated with osmolality due to it being an outlier (< 1 standard deviation below the mean of all samples).

Statistical analyses

Trial initiation times

Trial initiation time was defined as the duration of time in between the termination of one trial (violation, opt out, or the last lick from the rewarded port) and the subsequent center poke that began the next trial. We excluded outlier trial initiation times that were above the 99th percentile of the rat's cumulative trial initiation time distribution pooled over sessions. For all analyses, unless otherwise noted, the trial initiation time was associated with the identity of the trial it followed (block, reward volume). Sensitivity of trial initiation times to blocks was defined as change in initiation time (mean in low - mean in high blocks). Two-sided Wilcoxon signed-rank tests were used to compare within subject groups (fertile, non-fertile) and Kruskal–Wallis tests were used to compare sensitivity to blocks across groups (females to males).

Trial history effects

To assess trial initiation time sensitivity to previous offers, we regressed trial initiation time against previous rewards. We focused only on mixed blocks. Reward offers from a different block (e.g., a previous high block) were given NaN values. Then, we regressed against the previous nine reward offers, not including the current trial, along with a constant offset. We set the reward for violation, catch, and opt-out trials to zero, since rats do not receive a reward on these trials. We used Matlab's builtin regress function to perform the regression. With the coefficients, we found the first non-significant coefficient (coefficient whose 95% confidence interval contained zero), and set that coefficient and all following coefficients to zero. We fit an exponential decay function, as described previously[28], to the regression weights and compared the regression weights of each previous reward offer, the time constant (tau), and coefficients for one trial back for fertile and non-fertile stages with Wilcoxon sign-rank tests.

Photometry

Trials that followed violations were excluded from the data presented in Fig. 2. ΔAUC was defined as the difference between responses to offer cue in high - low blocks. For mixed blocks, the slope of AUC plotted as a function of reward volume was calculated for fertile and non-fertile stages. Two-sided Wilcoxon signed-rank tests compared ΔAUC and the slope of AUC as a function of reward volume in fertile to non-fertile stages.

Optogenetics

Only control sessions that occurred before the first day of optogenetic stimulation were included in the analysis because optogenetic activation of dopamine terminals had effects on behavior that lasted beyond the day of stimulation. Initiation times were normalized to high blocks in control sessions for both optogenetic and sham groups.

Protein quantification and localizing

Photomicrographs were analyzed in ImageJ. Mean fluorescent area was compared across groups with a Kruskal-Wallis test, followed by a Wilcoxon rank-sum test test for pair-wise comparisons if the group effect was significant. Relative proportions of membranous and intracellular DAT particles were compared in estrus and diestrus with Wilcoxon rank-sum tests. Their distance from the membrane was compared with a two-sample Kolmogorov-Smirnov test.

Significantly differentially expressed protein signatures identified with mass spectrometry were subjected to functional annotation. STRING database v12 was used to assess enrichment of gene ontology terms for rattus norvegus using a hypergeometric test and all identified proteins as background. According to Geneontology.org, 246 genes and their proteins are associated with dopamine. 92 of these proteins were identified by mass spectrometry and three that were significantly decreased in expression are related to dopamine uptake and transport (Slc6a3,

Slc6a4, and Smpd3).

Data Availability

The data generated in this study will be deposited in a Zenodo database upon publication.

Code Availability

Code used to acquire behavioral data, analyze all data, and generate figures will be available at `https://github.com/constantinoplelab/published/tree/main` upon publication.

References

1. McEwen, B. S. & Milner, T. A. Understanding the broad influence of sex hormones and sex differences in the brain. *Journal of neuroscience research* **95,** 24–39 (2017).
2. Handy, A. B., Greenfield, S. F., Yonkers, K. A. & Payne, L. A. Psychiatric symptoms across the menstrual cycle in adult women: a comprehensive review. *Harvard review of psychiatry* **30,** 100 (2022).
3. Riecher-Rössler, A., Butler, S. & Kulkarni, J. Sex and gender differences in schizophrenic psychoses—a critical review. *Archives of women's mental health* **21,** 627–648 (2018).
4. Burt, V. K., Altshuler, L. L. & Rasgon, N. Depressive symptoms in the perimenopause: prevalence, assessment, and guidelines for treatment. *Harvard Review of Psychiatry* **6,** 121–132 (1998).
5. Di Paolo, T. Modulation of brain dopamine transmission by sex steroids. *Reviews in the Neurosciences* **5,** 27–42 (1994).
6. Becker, J. B. Gender differences in dopaminergic function in striatum and nucleus accumbens. *Pharmacology Biochemistry and behavior* **64,** 803–812 (1999).
7. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275,** 1593–1599 (1997).
8. Waelti, P., Dickinson, A. & Schultz, W. Dopamine responses comply with basic assumptions of formal learning theory. *Nature* **412,** 43–48 (2001).
9. Bayer, H. M. & Glimcher, P. W. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* **47,** 129–141 (2005).
10. Cohen, J. Y., Haesler, S., Vong, L., Lowell, B. B. & Uchida, N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *nature* **482,** 85–88 (2012).
11. Day, J. J., Roitman, M. F., Wightman, R. M. & Carelli, R. M. Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nature neuroscience* **10,** 1020–1028 (2007).
12. Kim, H. R. *et al.* A unified framework for dopamine signals across timescales. *Cell* **183,** 1600–1616 (2020).
13. Steinberg, E. E. *et al.* A causal link between prediction errors, dopamine neurons and learning. *Nature neuroscience* **16,** 966–973 (2013).
14. Parker, N. F. *et al.* Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. *Nature neuroscience* **19,** 845–854 (2016).
15. Sharpe, M. J. *et al.* Dopamine transients are sufficient and necessary for acquisition of model-based associations. *Nature Neuroscience* **20,** 735–742 (2017).

16. Hamid, A. A. *et al.* Mesolimbic dopamine signals the value of work. *Nature neuroscience* **19,** 117–126 (2016).

17. Marino, M., Galluzzo, P. & Ascenzi, P. Estrogen signaling multiple pathways to impact gene transcription. *Current genomics* **7,** 497–508 (2006).

18. Vasudevan, N. & Pfaff, D. W. Non-genomic actions of estrogens and their interaction with genomic actions in the brain. *Frontiers in neuroendocrinology* **29,** 238–257 (2008).

19. Yoest, K. E., Quigley, J. A. & Becker, J. B. Rapid effects of ovarian hormones in dorsal striatum and nucleus accumbens. *Hormones and behavior* **104,** 119–129 (2018).

20. Cox, J. & Witten, I. B. Striatal circuits for reward learning and decision-making. *Nature Reviews Neuroscience* **20,** 482–494 (2019).

21. Floresco, S. B. The nucleus accumbens: an interface between cognition, emotion, and action. *Annual review of psychology* **66,** 25–52 (2015).

22. Sutton, R. S. & Barto, A. G. *Reinforcement learning: An introduction* (MIT press, 2018).

23. Joel, D., Niv, Y. & Ruppin, E. Actor–critic models of the basal ganglia: New anatomical and computational perspectives. *Neural networks* **15,** 535–547 (2002).

24. Chen, R. & Goldberg, J. H. Actor-critic reinforcement learning in the songbird. *Current opinion in neurobiology* **65,** 1–9 (2020).

25. Niv, Y., Daw, N. D., Joel, D. & Dayan, P. Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology* **191,** 507–520 (2007).

26. Xu-Wilson, M., Zee, D. S. & Shadmehr, R. The intrinsic value of visual information affects saccade velocities. *Experimental Brain Research* **196,** 475–481 (2009).

27. Wang, A. Y., Miura, K. & Uchida, N. The dorsomedial striatum encodes net expected return, critical for energizing performance vigor. *Nature neuroscience* **16,** 639–647 (2013).

28. Mah, A., Schiereck, S. S., Bossio, V. & Constantinople, C. M. Distinct value computations support rapid sequential decisions. *Nature Communications* **14,** 7573 (2023).

29. Niv, Y., Joel, D. & Dayan, P. A normative perspective on motivation. *Trends in cognitive sciences* **10,** 375–381 (2006).

30. Cox, J. *et al.* A neural substrate of sex-dependent modulation of motivation. *Nature Neuroscience,* 1–11 (2023).

31. Santollo, J., Edwards, A. A., Howell, J. A. & Myers, K. E. Bidirectional effects of estradiol on the control of water intake in female rats. *Hormones and behavior* **133,** 104996 (2021).

32. Barron, W. M., Schreiber, J. & Lindheimer, M. D. Effect of ovarian sex steroids on osmoregulation and vasopressin secretion in the rat. *American Journal of Physiology-Endocrinology and Metabolism* **250,** E352–E361 (1986).

33. Larsen, M. B. *et al.* Dopamine transport by the serotonin transporter: a mechanistically distinct mode of substrate translocation. *Journal of Neuroscience* **31,** 6605–6615 (2011).

34. Kim, S. K. *et al.* Neutral sphingomyelinase 2 induces dopamine uptake through regulation of intracellular calcium. *Cellular signalling* **22,** 865–870 (2010).

35. Watson, C. S. *et al.* Estradiol effects on the dopamine transporter–protein levels, subcellular location, and function. *Journal of molecular signaling* **1,** 1–14 (2006).

36. Fleming, W., Jewell, S., Engelhard, B., Witten, D. M. & Witten, I. B. Inferring spikes from calcium imaging in dopamine neurons. *PloS one* **16,** e0252345 (2021).

37. Atcherley, C. W., Wood, K. M., Parent, K. L., Hashemi, P. & Heien, M. L. The coaction of tonic and phasic dopamine dynamics. *Chemical Communications* **51,** 2235–2238 (2015).

38. Zhang, L., Doyon, W. M., Clark, J. J., Phillips, P. E. & Dani, J. A. Controls of tonic and phasic dopamine transmission in the dorsal and ventral striatum. *Molecular pharmacology* **76,** 396–404 (2009).

39. Beierholm, U. *et al.* Dopamine modulates reward-related vigor. *Neuropsychopharmacology* **38,** 1495–1503 (2013).

40. Cagniard, B., Balsam, P. D., Brunner, D. & Zhuang, X. Mice with chronically elevated dopamine exhibit enhanced motivation, but not learning, for a food reward. *Neuropsychopharmacology* **31,** 1362–1370 (2006).

41. Melikian, H. E. Neurotransmitter transporter trafficking: endocytosis, recycling, and regulation. *Pharmacology & therapeutics* **104,** 17–27 (2004).

42. Chen, R., Furman, C. A. & Gnegy, M. E. Dopamine transporter trafficking: rapid response on demand. *Future neurology* **5,** 123–134 (2010).

43. Miranda, M., Wu, C. C., Sorkina, T., Korstjens, D. R. & Sorkin, A. Enhanced ubiquitylation and accelerated degradation of the dopamine transporter mediated by protein kinase C. *Journal of Biological Chemistry* **280,** 35617–35624 (2005).

44. Vicentic, A., Battaglia, G., Carroll, F. I. & Kuhar, M. J. Serotonin transporter production and degradation rates: studies with RTI-76. *Brain research* **841,** 1–10 (1999).

45. Calipari, E. S. *et al.* Dopaminergic dynamics underlying sex-specific cocaine reward. *Nature communications* **8,** 13877 (2017).

46. Zhang, D., Yang, S., Yang, C., Jin, G. & Zhen, X. Estrogen regulates responses of dopamine neurons in the ventral tegmental area to cocaine. *Psychopharmacology* **199,** 625–635 (2008).

47. Shughrue, P. J., Lane, M. V. & Merchenthaler, I. Comparative distribution of estrogen receptor-α and-β mRNA in the rat central nervous system. *Journal of Comparative Neurology* **388,** 507–525 (1997).

48. Creutz, L. M. & Kritzer, M. F. Estrogen receptor-β immunoreactivity in the midbrain of adult rats: Regional, subregional, and cellular localization in the A10, A9, and A8 dopamine cell groups. *Journal of Comparative Neurology* **446,** 288–300 (2002).

49. McHenry, J. A. *et al.* Hormonal gain control of a medial preoptic area social reward circuit. *Nature neuroscience* **20,** 449–458 (2017).

50. Bocklisch, C. *et al.* Cocaine disinhibits dopamine neurons by potentiation of GABA transmission in the ventral tegmental area. *Science* **341,** 1521–1525 (2013).

51. Morales, M. & Margolis, E. B. Ventral tegmental area: cellular heterogeneity, connectivity and behaviour. *Nature Reviews Neuroscience* **18,** 73–85 (2017).

52. Matsuda, W. *et al.* Single nigrostriatal dopaminergic neurons form widely spread and highly dense axonal arborizations in the neostriatum. *Journal of Neuroscience* **29,** 444–453 (2009).

53. Zhu, J & Reith, M. Role of the dopamine transporter in the action of psychostimulants, nicotine, and other drugs of abuse. *CNS & Neurological Disorders-Drug Targets (Formerly Current Drug Targets-CNS & Neurological Disorders)* **7,** 393–409 (2008).

54. Sotnikova, T. D., Beaulieu, J.-M., Gainetdinov, R. R. & Caron, M. G. Molecular biology, pharmacology and functional role of the plasma membrane dopamine transporter. *CNS & Neurological Disorders-Drug Targets (Formerly Current Drug Targets-CNS & Neurological Disorders)* **5,** 45–56 (2006).

55. Redish, A. D. Addiction as a computational process gone awry. *Science* **306,** 1944–1947 (2004).

56. Montague, P. R., Dolan, R. J., Friston, K. J. & Dayan, P. Computational psychiatry. *Trends in cognitive sciences* **16,** 72–80 (2012).

57. Radulescu, A. & Niv, Y. State representation in mental illness. *Current opinion in neurobiology* **55,** 160–166 (2019).

58. Marcondes, F., Bianchi, F. & Tanno, A. Determination of the estrous cycle phases of rats: some helpful considerations. *Brazilian journal of biology* **62,** 609–614 (2002).

59. Creamer, M. S., Chen, K. S., Leifer, A. M. & Pillow, J. W. Correcting motion induced fluorescence artifacts in two-channel neural imaging. *PLoS computational biology* **18,** e1010421 (2022).

60. Santiago, A. N., Lim, K. Y., Opendak, M., Sullivan, R. M. & Aoki, C. Early life trauma increases threat response of peri-weaning rats, reduction of axo-somatic synapses formed by parvalbumin cells and perineuronal net in the basolateral nucleus of amygdala. *Journal of Comparative Neurology* **526,** 2647–2664 (2018).

61. Chen, Y.-W., Wable, G. S., Chowdhury, T. G. & Aoki, C. Enlargement of axo-somatic contacts formed by GAD-immunoreactive axon terminals onto layer V pyramidal neurons in the medial prefrontal cortex of adolescent female mice is associated with suppression of food restriction-evoked hyperactivity and resilience to activity-based anorexia. *Cerebral cortex* **26,** 2574–2589 (2016).

62. Spiteri, T., Ogawa, S., Musatov, S., Pfaff, D. W. & Ågmo, A. The role of the estrogen receptor α in the medial preoptic area in sexual incentive motivation, proceptivity and receptivity, anxiety, and wheel running in female rats. *Behavioural brain research* **230,** 11–20 (2012).

63. El-Neweshy, M. S. Experimental doxycycline overdose in rats causes cardiomyopathy. *International Journal of Experimental Pathology* **94,** 109–114 (2013).

Acknowledgments

We thank members of the Constantinople lab for feedback. We thank research technicians in the Constantinople lab for animal training. We thank Claudia Farb for her assistance with histology and electron microscopy and NYU Langone's Proteomics Laboratory (SCR_017926) for the mass spectrometry. We thank Ina Baek Bando and Gina Canino for their assistance with histology processing, imaging, and quantifying.

Funding: This work was supported by a K99/R00 Pathway to Independence Award (R00MH111926), a Klingenstein-Simons Fellowship in Neuroscience, and an NIH Director's New Innovator Award (DP2MH126376) to C.M.C. C.G. was supported by a grant from the Simons Foundation (855332)), F32MH125448, and 5T32MH019524. The mass spectrometric experiments were supported in part by NYU Langone Health, the Laura and Isaac Perlmutter Cancer Center support grant P30CA016087 from the National Cancer Institute, and the NIH Shared Instrumentation Grant 1S10OD010582-01A1 for the purchase of an Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer.

Author Contributions

C.M.C. and C.E.G. designed the study. C.E.G., D.K., A.C.M. and D.H.L. collected and analyzed data. A.M. developed the behavioral model and contributed to data analysis. T.Y. and D.L. created and shared the shEsr1 virus. C.A. supervised the electron microscopy. C.E.G prepared the figures. C.E.G. and C.M.C. wrote the manuscript. C.M.C supervised the project.