Chronic Ethanol Exposure Produces Sex-Dependent Impairments in Value Computations in the Striatum

Yifeng Cheng[1,5]*, Robin Magnard[1], Angela J. Langdon[2], Daeyeol Lee[1,3,4,5], and Patricia H. Janak[1,4,5]*

[1]Department Psychological and Brain Sciences, Krieger School of Arts and Sciences, Johns Hopkins University, Baltimore, MD;
[2]Intramural Research Program, National Institute of Mental Health, National Institutes of Health, Bethesda, MD;
[3]Zanvyl Krieger Mind/Brain Institute, Krieger School of Arts and Sciences, Johns Hopkins University, Baltimore, MD;
[4]Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD;
[5]Kavli Neuroscience Discovery Institute, Johns Hopkins University, Baltimore, MD.

*Corresponding authors (patricia.janak@jhu.edu; ycheng62@jhu.edu)

Value-based decision-making relies on the striatum, where neural plasticity can be altered by chronic ethanol (EtOH) exposure, but the effects of such plasticity on striatal neural dynamics during decision-making remain unclear. This study investigated the long-term impacts of EtOH on reward-driven decision-making and striatal neurocomputations in male and female rats using a dynamic probabilistic reversal learning task. Following a prolonged withdrawal period, EtOH-exposed male rats exhibited deficits in adaptability and exploratory behavior, with a preference for value updating based on rewards rather than omissions. These behavioral changes were linked to altered neural encoding in the dorsomedial striatum (DMS), where EtOH increased outcome-related signals and decreased choice-related signals. In contrast, female rats showed minimal behavioral changes with distinct EtOH-evoked alterations of neural signals, revealing significant sex differences in the impact of chronic EtOH. Our findings underscore the profound impact of chronic EtOH exposure on adaptive decision-making, revealing enduring changes in neurocomputational processes in the striatum underlying cognitive deficits that differ by sex.

Introduction

Value-based decision-making is a fundamental cognitive process that involves evaluating the potential outcomes of different actions to choose the most beneficial one. This process is crucial for adaptive behavior and relies heavily on neural computations in the striatum, a key brain region for encoding various decision-related signals, including those related to choice, rewards, and value (1-9).

Chronic ethanol (EtOH) exposure is known to enduringly disrupt learning and cognition in humans (10-13) and rodents (14-16), and profoundly affect the brain, particularly in regions involved in decision-making and reward processing, such as corticostriatal pathways. Previous research has demonstrated that EtOH alters striatal cell signaling and corticostriatal plasticity (17-20), especially in the dorsomedial striatum (DMS) (18-20), a region critical for economic decision-making (4, 9, 21, 22), reward learning (1-3, 8, 23, 24), and goal-directed action

planning (25, 26). However, how chronic alcohol exposure affects specific decision processes underlying reward learning and disrupts systems-level striatal dynamics of these decision processes remains underexplored.

Many psychiatric diseases, including alcohol use disorders, disproportionately affect one sex over the other, underscoring the importance of considering sex differences in research (27). Studies indicate that males and females exhibit distinct patterns of alcohol consumption and susceptibility to alcohol-related disorders (28), potentially driven by differences in neurobiology (29, 30). Moreover, males and females have been shown to exhibit distinct behavioral responses and motivational levels in reward learning (5, 31). These sex differences may extend to the effects of alcohol on decision-making and the underlying neural mechanisms. Thus, it is crucial to understand how chronic ethanol exposure influences striatal encoding of reward learning signals in a sex-dependent manner, as these neural changes can lead to persistent impairments in decision-making.

To address these questions, we trained both male and female rats that had undergone a period EtOH-dependence in a dynamic probabilistic reversal learning (dynaPRL) task. The dynaPRL task requires rats to assess reward probabilities under varying levels of uncertainty, enabling a rigorous examination of decision-making processes. Using this task, we observed EtOH-induced changes across multiple behavioral measures, including altered exploration-exploitation trade-offs and reversal deficits in more uncertain environments, which were more pronounced in male rats than in female rats. Using a reinforcement learning framework, we found that both EtOH-exposed male and female rats showed enhanced value updating after rewards, but only males exhibited specific alterations after no reward, offering a potential computational explanation for the observed deficits. Additionally, single-unit recordings in the DMS revealed reduced encoding of choice, and state and policy value, alongside an enhanced encoding of outcome and chosen value in male rats with an EtOH history. Notably, our observations revealed that female rats displayed a distinct pattern of encoding for decision variables compared to males, irrespective of EtOH history, and EtOH history altered neural signal encoding in females differently than it did in males, underscoring significant sex-specific effects. Together, these findings reveal behavioral and neural mechanisms underlying effects of chronic EtOH exposure during protracted withdrawal, highlighting sex differences in striatal processing of decision-making signals, indicating that the impact of EtOH on value-based decisions can vary markedly between male and female rats.

Results

Protracted withdrawal from chronic EtOH exposure produces a sex-specific behavioral pattern

To investigate whether chronic EtOH exposure disrupts decision-making beyond the phase of acute withdrawal (i.e., >1 week), we used a well-validated chronic intermittent EtOH (CIE) vapor procedure to model EtOH dependence (blood EtOH concentration, BEC > 150 mg/dl) in rats (**Figure 1A**)(14, 32). In our initial experiment, rats were trained on a standard probabilistic reversal learning (PRL) task and then underwent cycles of EtOH vapor or air exposure and withdrawal for four weeks (16 hrs/day, 5 days/wk), with group assignment balanced for PRL performance. Following 10-15 days of withdrawal, performance on the standard PRL task was reassessed (**Figure S1A**). This task required water-restricted rats to earn a drop (33 µl) of 10% sucrose water by choosing between left and right levers, which delivered a reward with either a 70% or a 10% probability, while the reward probabilities repeatedly switched

within a session (**Figure 1B and S1B**). Consistent with previous studies (33, 34), during this standard PRL task, we did not observe any significant differences in reversal learning after block switches when comparing within-subject performance before and after the CIE exposure (**Figures S1C-1G**), nor between EtOH and air controls in either male or female rats (**Figures S1C-1G**). To further understand whether the CIE exposure could induce behavioral changes in this task, we employed a Support Vector Machine analysis (SVM) to decode group membership from a high-dimensional behavioral dataset consisting of 20 behavioral measures gathered during pre- and post-reversal phases in EtOH rats and their air controls (**Figure S2A**). We found that trained SVM models could not identify either group labels or sex labels in this PRL task above chance levels (**Figure S2B**). These findings indicate that rats did not express significant EtOH-induced performance changes within the standard PRL task.

To further explore possible effects of chronic EtOH on cognitive flexibility, we next modified the task to increase the level of uncertainty (35, 36). We introduced a three-block design with reward probabilities that varied in the contrast between left and right lever probabilities (80:10, 60:30, and 45:45; **Figure 1B and 1C**), converting the standard PRL task into a dynamic PRL task (dynaPRL). Consequently, when blocks transitioned from one to another, rats faced varying amounts of uncertainty such that a larger change in the reward probabilities would be easier to detect and might facilitate adaptive changes in behavior (**Figure 1C**). Individual rats in both air and EtOH groups adjusted their choice behavior in response to block transitions within the dynaPRL task (**Figure 1D**). Applying the SVM model to the same behavioral features used in the analysis for the standard PRL task, we found the model could correctly identify group (air vs. EtOH) and sex labels from the dynaPRL behavioral data with a prediction precision (63.29%) significantly higher than chance levels (25%; $p < 0.001$). The trained SVM decoder almost never (<0.001% probability) misclassified male and female subjects (**Figures 1E, 1F and S2C-S2G**). Moreover, the decoding accuracy for EtOH treatment vs. air was significantly higher than chance for male but not female rats (**Figure 1F**). Using uniform Manifold Approximation Projection (UMAP), we found that the data were distinctly segregated for male EtOH and air rats in a 2D low-dimensional space with an accuracy (~70%), similar to the SVM decoding of group membership (**Figure 1G**).

Taken together, these results indicate that the impact of prior chronic EtOH exposure can be detected in male rats many weeks (>10 wks) after the last EtOH exposure through multivariate analysis of high dimensional behavioral measures during the more challenging dynaPRL decision-making task.

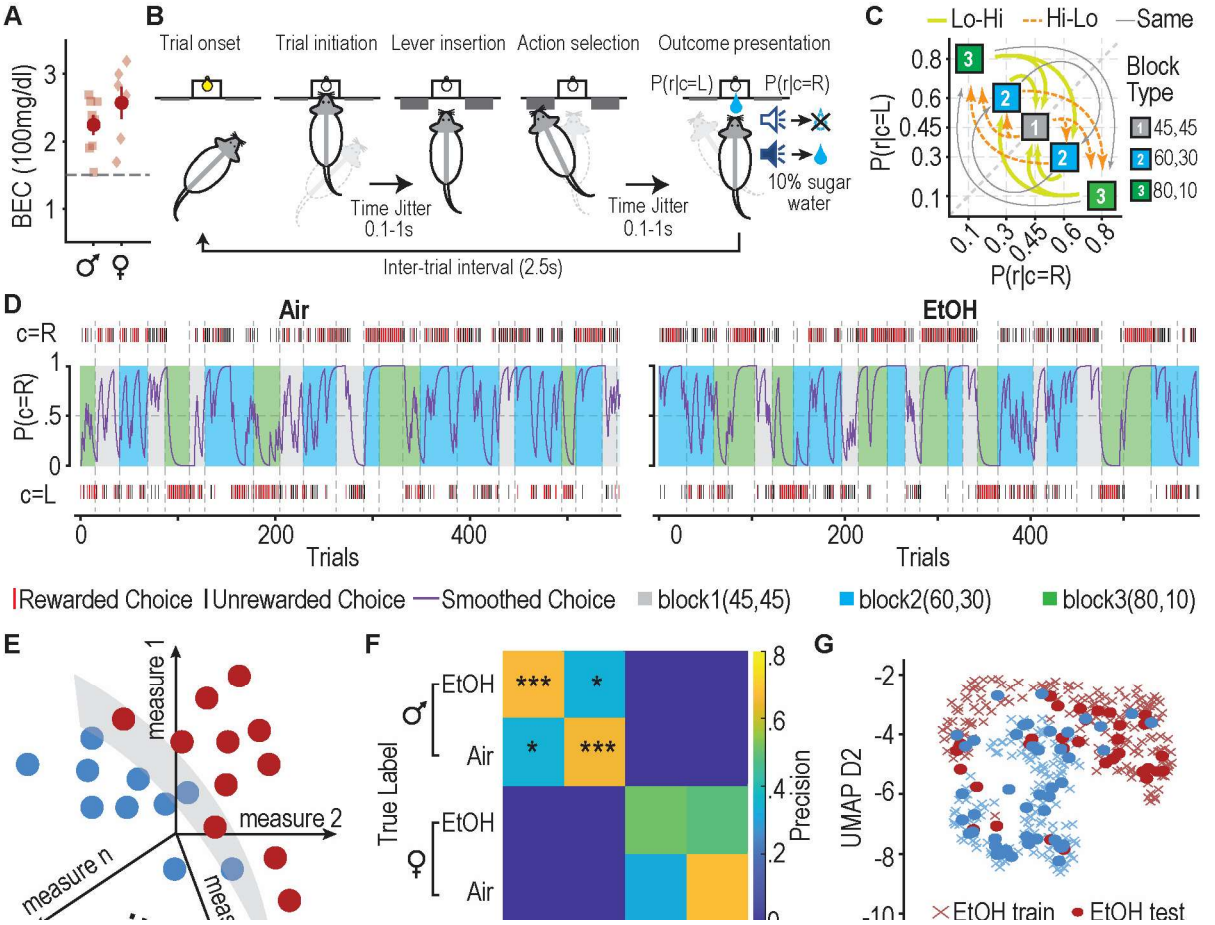

Figure 1. EtOH-exposed rats exhibit a distinct pattern of behavior in the dynamic probabilistic reversal learning task (dynaPRL) after prolonged withdrawal. (**A**) Blood ethanol (EtOH) concentrations (BEC) of male and female rats during EtOH vapor exposure. $t(12) = -1.24$, $p = 0.239$, t test. (**B**) Schematic of trial structure illustrating each behavioral stage in the dynaPRL task. Trials begin with illumination of the magazine light (trial onset). Rats can then initiate a trial with a magazine entry (trial initiation) that is followed by insertion of both levers (lever insertion). Once rats press one of the levers (action selection), levers are withdrawn, and an outcome is presented (outcome presentation). A rewarded choice is indicated by a clicker sound, followed by sucrose, and an unrewarded choice is indicated by white noise. P(r|c=L) or p(r|c=R is the reward (r) probability given that the choice (c) is left (L) or right (R) side. (**C**) Choice reward probabilities for the dynaPRL task. The reward probabilities associated with each lever switched among 3 distinct blocks, drawing probabilities from three different sets, (0.8, 0.1), (0.6,0.3), and (0.45, 0.45). Based on the difference of reward probabilities between two levers before and after block transition, all block transitions can be categorized into three types: low to high (LH) challenge, high to low (HL) challenge, and same to same (no difference, ND), denoted as lines with distinct colors, styles and thickness. (**D**) Two example rats from air (left) and EtOH (right) groups exhibit choice behavior that changes according to block switching. Choices are depicted with red hashmarks for rewarded choice, grey for unrewarded. The letter 'c' denotes the side of choice, with 'L' for left and 'R' for right. A smoothed estimate of the probability of choosing the right side, P(c=R), is calculated using a moving average across a 5-trial window. (**E**) Schematic illustrating how a hyperplane, depicted by the grey surface, differentiates between sessions from EtOH-exposed rats and air-exposed rats, represented as dots, within the multidimensional space created by various behavioral measurements. This illustration represents our hypothesis that EtOH-exposed and air-exposed groups establish distinct behavioral patterns. (**F**) Confusion matrix for the SVM multiclass classifier

aggregating from 1000 iterations for decoding both group and sex for n = 13 EtOH rats (8 male, 5 female) and 14 air controls (9 male, 5 female). Entries on the primary diagonal represent accurate predictions, off-diagonal elements represent misclassifications. Classification precision calculated based on true positive rate/(true positive rate + false positive rate). *$p < 0.05$, ***$p < 0.001$, Monte Carlo significance tests. (**G**) A UMAP transformation of high dimensional data onto a 2D space.

Chronic EtOH exposure slows reversal learning and reduces exploration only in males

To further dissect EtOH-induced alterations in the dynaPRL task, we examined how EtOH and air rats adjusted their choice behaviors in response to changes in reward probabilities after a block switch. We specifically analyzed the probability of selecting the action that was advantageous before the reversal. When a block switched from low uncertainty to high uncertainty (Lo-Hi) (**Figure 2A**), a two-way mixed-effect ANOVA revealed a significant interaction between treatment group and trials in males (**Figure 2B**; $p = 0.016$). To quantify reversal learning performance in male rats, we fit a single-exponential decay model to the probability of choosing the previously preferred action across the first 12 trials after the block transition (**Figure 2C**). While there was no difference in pre-reversal asymptotic value (*a)* or the magnitude of change in learning new action-outcome contingency after the reversal (*d*) (**Figures 2D and 2E**), we observed a lower adaptation rate (λ) in EtOH males compared to the same-sex controls (**Figure 2F**; $p = 0.016$). Given that the inverse of λ ($\tau = 1/\lambda$) corresponds to the time constant of the rat's learning in a new block, this indicates that EtOH rats needed more experience to adapt to a new contingency in a highly ambiguous environment (**Figure 2G**; $p = 0.036$). Notably, these effects were not observed during block transitions from high to low uncertainty (Hi-Lo), at the same uncertainty level before and after block transitions (same), or when all trial types were combined (**Figures S3B-3D**). By contrast, in females, we found that female EtOH and air rats exhibit identical reversal performance across all transition types (**Figure S3G-SJ**). Thus, consistent with the results from the SVM analysis, the impact of EtOH on reversal learning dynamics was present only in male rats when they face a sudden transition to high uncertainty. To examine choice behavior later in a block, after rats are expected to have acquired the new contingencies, we focused on behavior after 12 trials from block switch and assessed how much choice probability deviated from local reward rate yielded by the choice ('matching behavior') (37). Male, but not female, EtOH rats aligned their choices with local reward rates more tightly than same-sex control rats (**Figure S4**).

To investigate whether and how EtOH might differentially impact choice strategies in male and female rats, we analyzed their stay probabilities in two phases, the early reversal learning phase, including trials 2 to 7 after transitioning to a new block, and the late performance phase of the current block, spanning the last 6 trials before block switch. We found that male EtOH rats were more likely to repeat the same choice in two consecutive trial, especially in the early phase following a switch, but less in the late phase after rats have adapted to the contingency change, as shown by measures of stay probabilities and the differences between win-stay and lose-shift probabilities (**Figures 2H-2J and S5A-S5H**). This behavior was sensitive to the receipt or omission of reward; EtOH rats exhibited higher stay probabilities after reward omission, but not reward receipt, during 'Lo-Hi' transitions (**Figure 2I**), resulting in a greater tendency for win-stay behavior and less

tendency for lose-shift behavior (**Figure 2J**). By contrast, female EtOH exposed rats did not show alterations in win-stay lose-shift behavior strategies (**Figures S5I-S5P**). Notably, EtOH-naive male rats displayed higher stay probabilities and a greater tendency for win-stay behavior compared to EtOH-naive female rats (Figures S2F).

Collectively, these results demonstrate that chronic EtOH exposure significantly impacts outcome-driven action strategies, favoring a shift from switch to stay and enhancing matching behaviors in male, but not female, rats, i.e., a tendency for greater exploitation and less exploration, resulting in a sex-dependent reduction in cognitive flexibility during a dynamic decision-making task.

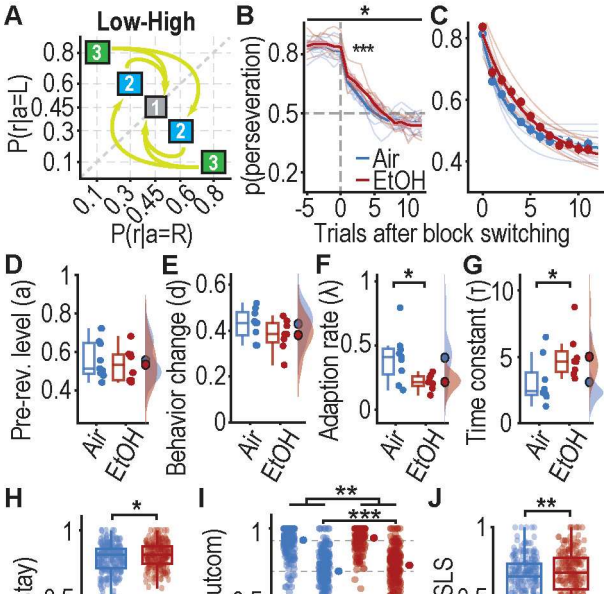

Figure 2. Protracted withdrawal from chronic EtOH exposure alters choice behaviors of male rats in highly ambiguous environments. (**A**) The diagram illustrates block switching in Low-to-High uncertainty transitions. **(B)** Probability of choosing the pre-switch preferred lever in Low-High type transitions. Bold lines represent group averages and faded lines represent individual rats. Vertical dashed lines indicate the last trial before block switching (t=0). Main effect of group, $F(1, 16) = 1.708$, $p = 0.211$; main effect of time, $F(16,255) = 362.055$, ***p = 1.2E-167; group x time interaction, $F(16,255) = 1.925$, *p = 0.016, two-way mixed-effect ANOVA. *p < 0.05 for post hoc test with Bonferroni correction. (**C**) A single exponential decay model estimates performance in learning new action-outcome contingency. Bold lines show the average model fit for probability of selecting an action based on the pre-switch contingencies. Dots depict the true mean data from B. Individual rat model estimates of choice probability are shown in faded lines. **(D-G)** The estimated parameters' pre-reversal asymptotic value (pre-rev. level), a (D); magnitude of change in p(preservation) (behavior change), d (E); and adaptation rate, λ (F) for the single exponential decay model in C. The inverse of λ, τ, represents the time constant of learning in a new block (G). The box plot depicts the median, 25th, and 75th percentile of parameter values for each group. Dots next to the box plot values from individual rats. Cloud plots on the right side of panels depict the data distribution. Dots on cloud plots are the mean of parameter values. $t(15) = -0.498$, $p = 0.626$ for D; $t(15) = -1.522$, $p = 0.149$ for E; $t(15) = -2.704$, *p = 0.016 for F; $t(15) = 2.300$, *p = 0.036 for G, two-sample independent t test with two-tails. n = 9 Air rats, 8 CIE rats

for B-G. (**H**) Stay probability, the likelihood of repeating the same action on the current trial as the last trial, during 'Lo-Hi' transitions. z(464) = -2.358, *p = 0.018, Wilcoxon rank-sum test. (**I**) Stay probability given reward (water drop icon) or no reward (cross, X, symbol) on the prior trial, during 'Lo-Hi' transitions. Dots next to scatter plots represent the data means. Standard errors are vertical lines but are too small to visualize on the graph. Horizontal dashed lines represent the means of air group. Main effect of outcome, F(1, 463) = 1076.3, ***p = 7.3E-123; main effect of group, F(1, 463) = 10.625, **p = 0.001; outcome x group interaction, F(1, 463) = 4.5, *p = 0.035, two-way mixed-effect ANOVA. ***p < 0.001 post hoc test with Bonferroni correction. (**J**) Difference (Δ) between probability of stay given reward (win-stay, WS) and probability of shift given no reward (lose-shift, LS), ΔWSLS, during 'Lo-Hi' transitions. z(464) = -2.805, **p = 0.005, Wilcoxon rank-sum test. N = 241 sessions from 9 air rats, 229 sessions from 8 EtOH rats for H-J.

Chronic EtOH exposure alters outcome-specific value updating in male rats

To better understand the reward learning processes underlying the observed effects of EtOH exposure on decision-making, we utilized computational tools by applying several reinforcement learning (RL) models to the trial-by-trial choice and reward data. We found that a Q-learning model with differential learning rates (model 2, Q-DFLr) outperformed all other candidate models (**Figure S6A**). Q-DFLr has four distinct learning parameters controlling the updating of the value function: α^+ and α^- for updating the value of chosen actions when the outcome is reward or no reward, respectively, and θ^+ and θ^- for discounting the value of unchosen actions in an outcome-dependent fashion (**Figure 3A**). To assess the group-level differences between the EtOH and air control groups across RL parameters, we constructed a hierarchical model wherein the distributions of individual parameters for each rat are determined by the group-level hyperparameters. These hyperparameters encompass the means of the distributions for each parameter, the mean differences (δ) between air and EtOH groups for each parameter, and the variances of each parameter's distribution. The posterior density of these hyperparameters was estimated through sampling using Hamiltonian Markov Chain Monte Carlo (HMC) methods. We observed a rightward shift in the posterior distributions of the group difference parameter, δ, for the learning rate α^+ in both male and female rats (**Figure 3B**). Using directed Bayes Factor (dBF) to assess the strength of this shift, we found that the rewarded learning rate changed relatively similarly across the sexes, with α^+ being approximately 6 times more likely to increase than decrease in males, and 5 times more likely in females, following EtOH pre-exposure compared to same-sex controls (**Figure 3B**; dBF = 6.1 for male and 5.5 for female). Notably, the unrewarded learning rate, α^-, was more likely to increase following EtOH exposure only in females (**Figure S6B**; dBF = 2.18), but not in males (**Figure S6B**; dBF < 1). Additionally, we found that the discounting parameter, θ^-, exhibited a pronounced rightward shift in males, where it was approximately 8 times more likely to increase than decrease after EtOH pre-exposure (**Figure 3C**; dBF = 8.35). In contrast, the likelihood of θ^- increasing after EtOH exposure was considerably lower in females (**Figure 3C**; dBF = 1.54). All other parameters showed only small differences between the two groups for both sexes (**Figure S6B**; all dBFs < 1). We then used the best-fit parameters to simulate choice behavior within the original block structure used in the experiment and analyzed the probabilities of perseveration for each block transition type.

We found that the simulated data partially recapitulated the EtOH-induced reversal deficits observed in the empirical data from male, but not female, rats when transitioning from a low to a high uncertainty block, but not for other types of block transitions (**Figures 3D and S6C**). Collectively, these data suggest that chronic EtOH exposure selectively alters reward learning in male, but not female, by enhancing both outcome-based updating of selected actions and omission-based forgetting of unselected options, biasing action-selection towards previously reinforced choices. These results provide an algorithmic explanation for the sex-dependent changes in choice behavior we observed.

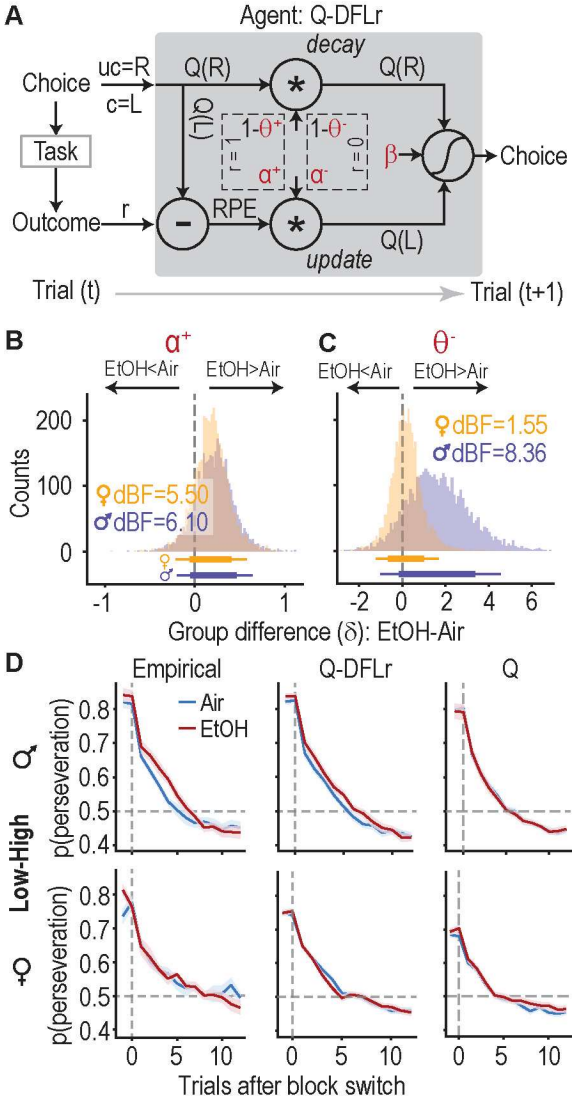

Figure 3. EtOH exposure enhances positive feedback-mediated learning rates and negative feedback-mediated forgetting rates. (**A**) Schematic representation of the value updating processes employed by the best-fitting RL agent, Q-DFLr. In this example, an agent makes a left side choice at trial t, denoted as 'c=L', where 'c' stands for 'choice'. 'uc=R' indicates the right lever was not selected, where 'uc' stands for 'unchosen'. Upon making a left side choice, the task generates an outcome, 'r', where '1' signifies reward, i.e., positive feedback, and '0' signifies absence of reward, i.e., negative feedback. The left action value, Q(L)—or the chosen value—is updated based on the reward prediction

error (RPE) using two distinct learning rates, α^+ for reward (r=1) and α^- for non-reward (r=0) action outcomes. Conversely, the right action value, Q(R)—or the unchosen value—undergoes decay at two distinct forgetting rates, θ^+ and θ^-, depending on the action outcome, r. The updated action values are then adjusted by the inverse temperature parameter, β, to inform the agent's next choice (t+1) via a softmax selection function. (**B,C**) Hamiltonian Markov chain Monte Carlo (HMC)-sampled posterior densities of group difference, δ, between EtOH- and air-exposed male (♂) and female (♀) rats in positive feedback-mediated learning rates (α^+) in (B) and negative feedback-mediated forgetting rates (θ^-) in (C), across group-level hyperparameters. The leftward arrow indicates that the parameter value is higher in EtOH compared to air, while the rightward arrow indicates that the parameter value is lower. Bottom horizontal lines represent the 80% and 95% highest density interval (HDI). dBF, directed Bayes factor. Posterior densities aggregated from 1000 times Monte Carlo sampling. (**D**) Choice probability of the initially preferred lever in low-to-high transitions. Data in left panels are empirical (reproduced from Figure 1G for male, Figure S4G for female), while the middle and right panels depict simulated data, generated using the best-fitting parameters from Q-DFLr and Q models, respectively. Each line represents an average of 20 simulations.

Sex-dependent effects of EtOH exposure on DMS activity related to actions and outcomes

Our behavioral analyses and reinforcement learning model fitting indicated that male, but not female, EtOH and air-control rats exhibit distinct outcome-dependent action value updating. This suggests that EtOH might selectively alter neural processes related to action and outcome signals in males. Therefore, we sought to examine the impact of EtOH on DMS neural dynamics within the dynaPRL task. We recorded 494 and 241 single units in the DMS from male EtOH and air control rats, respectively, and 354 and 347 single units from female EtOH and air control rats, respectively (**Figure S7A-S7C**).

Using multiple linear regression models, we examined how neural activity in different time epochs throughout the trial was affected by an animal's action, the outcome of the action, and the interaction of the selected action and its outcome, for both the current trial (t) and the previous trial (t-1) and compared the significance of this modulation at the group level using a generalized linear mixed-effects models (GLMM). Choice direction (ipsi- or contralateral to the recorded hemisphere) signals reflecting the action made on the current trial (**Figures 4A and 4B**), but not the previous trial (**Figures S8A and S8B**), were apparent in both groups and sexes throughout much of the trial. In both EtOH-exposed and control rats, a significant increase in choice-modulated neurons was observed from the intertrial interval to outcome presentation, with no sex-specific differences. Following onset of outcome, female rats showed significantly higher choice encoding than males, regardless of EtOH history ($p = 0.0022$ for main effect of sex). Additionally, a significant interaction between sex and EtOH treatment on neural encoding of choice direction during the outcome period was found ($p = 0.0096$), accounted for by fewer choice-modulated neurons in EtOH-exposed males relative to air controls ($p = 0.027$ for main effect of treatment), but more choice-modulated neurons in EtOH-exposed females than air controls, especially early in the outcome period ($p = 0.0096$ for interaction between treatment and time).

When examining the impact of current and prior trial outcomes on neural activity, we identified further sex-dependent differences in EtOH and air groups during the outcome period ($p = 0.013$, interaction among treatment, sex and time). In EtOH males, a higher proportion of DMS neurons (~80%) was modulated by the outcome compared to air controls (~50%) (**Figures 4C and 4D**; $p < 0.0001$). By contrast, in females, both EtOH

and air groups recruited similar proportions of neurons (~75%) immediately after outcome onset, but the signal diminished more slowly in EtOH females ($p = 0.0040$, interaction between treatment and time). To further dissociate this EtOH-driven effect in both sexes, we analyzed neuronal firing rates following rewarded and unrewarded outcomes and found a greater proportion of neurons modulated by both outcome types in EtOH males compared to controls (**Figures S9A and S9B**; chi-squared test, $p < 0.0001$), while females showed no significant treatment differences (**Figures S9C and S9D**). We also analyzed the impact of the outcome received from the previous trial on current trial neural firing, and found the prior reward was significantly encoded throughout the current trial, with pronounced sex-dependent differences observed during the ITI period in EtOH versus control groups ($p = 0.0069$) (**Figures S8C and S8D**). Specifically, previous-trial outcome signals diminished faster in EtOH males and slower in EtOH females compared to their respective controls ($p < 0.0001$ for interaction between treatment and time in both groups).

Lastly, we found that EtOH differentially impacted neural encoding dynamics of the action-outcome (A-O) interaction (a proxy for contingencies) in DMS neurons recorded from males compared to females (**Figures 4E and 4F**; $p = 0.00098$ for the interaction among treatment, sex, and time). The encoding of the action-outcome interaction was more prominent at the start of the outcome presentation in EtOH males, whereas it became more pronounced toward the end in air controls ($p = 0.044$ for interaction between treatment and time). By contrast, no significant differences were observed between neural populations recorded from EtOH and controls in females. Notably, this neural encoding of A-O was only detected when analyzing current trial events (**Figures S8E and S8F**).

Collectively, these results demonstrate that EtOH pre-exposure differentially modulates neural encoding of choice direction and outcome signals during outcome presentation in a sex-dependent manner. Additionally, EtOH impacts the dynamics of action-outcome interaction encoding in males, suggesting that EtOH alters the temporal processing of action-outcome contingencies in these subjects.

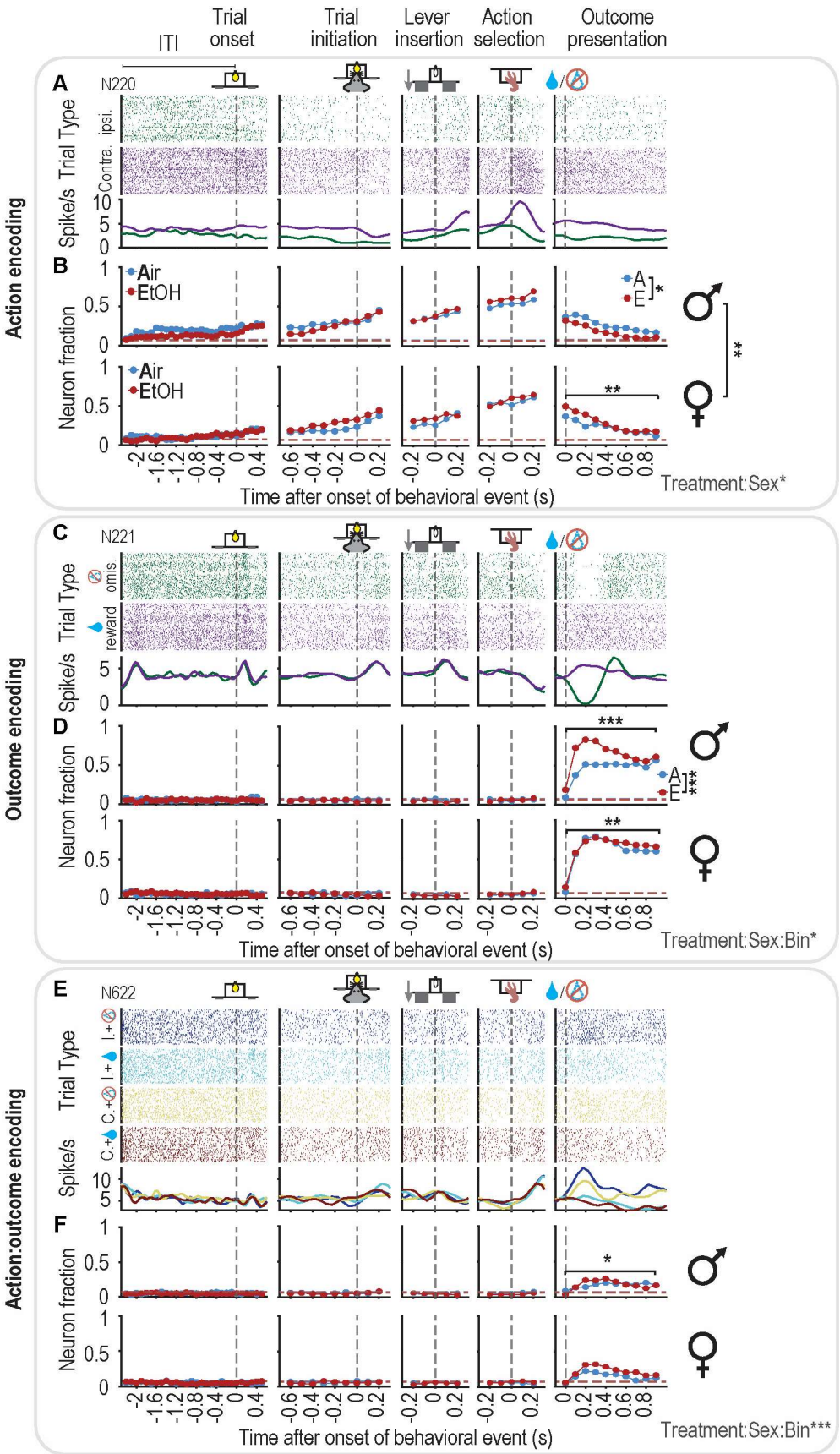
ITI
Trial onset
Trial initiation
Lever insertion
Action selection
Outcome presentation
Action encoding
A
N220
Trial Type
ipsi.
Contra.
Spike/s
B
Air
EtOH
Neuron fraction
Time after onset of behavioral event (s)
Treatment:Sex*
Outcome encoding
C
N221
omis.
reward
D
Treatment:Sex:Bin*
Action:outcome encoding
E
N622
I.+
C.+
F
Treatment:Sex:Bin***

Figure 4. **EtOH exposure alters encoding of action and outcome during outcome presentation.** (**A**) Example DMS unit from an air rat shows spike activity correlated with action side in the current trial. At top are spike raster plots, trials are grouped by choice side as either ipsilateral (ipsi) or contralateral (contra) to the recording site. (**B**) Fraction of neurons with activity significantly modulated by the current trial action across all trial landmarks, including intertrial interval (ITI) prior to trial onset, trial initiation, lever insertion, action selection, and outcome presentation. During outcome presentation, *p = 0.027 main effect of treatment in males; **p = 0.0096 interaction between treatment and time in females; **p = 0.0022 main effect of sex, *p = 0.032 interaction between treatment and sex. (**C**) Example DMS unit from the air group shows spike activity correlated with outcome (reward or omission) in the current trial. At the top are spike raster plots, trials are grouped by outcome as rewarded (reward) or unrewarded. (**D**) Fraction of neurons with activity significantly modulated by the current trial's outcome across all trial landmarks. During outcome presentation, ***p < 0.0001 interaction between treatment and time bin and ***p < 0.0001 main effect of sex in males; **p = 0.004 interaction between treatment and sex in females; *p = 0.013 interaction among treatment, sex and time bin. (**E**) Example DMS unit from air group shows spike activity correlated with interaction of choice and outcome (labeled action:outcome) in the current trial. At the top are spike raster plots, trials are grouped by distinct contingencies as ipsilateral choice without reward ('I' and no water drop icon), ipsilateral choice with reward ('I' and water drop icon), contralateral choice without reward ('C' and no water drop icon), and contralateral choice with reward ('C' and water drop icon). (**F**) Fraction of neurons with activity significantly modulated by the current trial's action:outcome interaction all trial landmarks. During outcome presentation, *p = 0.043 interaction between treatment and time bin in males; ***p = 0.00098 interaction among treatment, sex and time. At the bottom of A, C, and E are spike density functions generated by applying a Gaussian kernel (δ = 100 ms) to the corresponding spike trains. In the B, D, and F, the horizontal dashed lines indicate the chance level (binomial test, α = 0.05). Logistic regression with Generalized Linear Mixed-Effect Model for B,D and F. n = 241 neurons from 9 air male rats, 494 neurons from 8 EtOH male rats, 347 neurons from 5 air female rats, and 354 neurons from 5 EtOH female rats.

Chronic EtOH exposure persistently alters value signals in the DMS in a sex-dependent manner

Next, we evaluated whether EtOH impacted value signal encoding in the DMS during the action selection process, defined as the time period from the beginning of ITI until lever press. Specifically, we sought to quantify the strength and time course of striatal activity related to two orthogonal combinations of the action values (38) estimated by the best-fitting reinforcement learning model (model 2, Q-DFLr). The first variable is the summation of the two available action values, ΣQ, which represents the richness of the recent environment and roughly corresponds to the so-called state value. State value is likely to be important environmental information for inferring the block type. Although each block has the same average reward probabilities, the local reward rate reflects the uncertainty level of the block (**Figure S10**). Thus, a better representation of state value might aid rats in inferring their current environment. The second variable corresponds to the difference between the two available action values, ΔQ, which is monotonically related to the probability of selecting one action over the alternative and is referred to as the policy signal. We analyzed neural firing rates, aligned with either trial onset (including ITI), trial initiation, lever insertion, or lever press, as linear functions of ΔQ and ΣQ, while accounting for the effects of choice and chosen action value in the regression analysis.

We found that ~40% of striatal neurons across all groups and sexes exhibited changes in spike rate in response to the ΣQ (state value) at the beginning of the ITI, which gradually diminished to chance levels (binomial test, α =

0.05) around the lever insertion period (**Figures 5A, 5D, and 5G**). Notably, during the ITI period, EtOH exposure significantly modulated the neural encoding dynamics of state value in a sex-dependent manner ($p < 0.0001$ for the interaction among treatment, sex, and time). In males, the proportion of neurons encoding state value decreased more rapidly in EtOH-exposed rats compared to controls ($p < 0.0001$, **Figure 5D**), while no significant differences were observed in females (**Figure 5G**). Moreover, we observed approximately 20% of striatal neurons encoded action policy (ΔQ) at the beginning of the ITI across all groups and sexes (**Figures 5B, 5E, and 5H**). Throughout the ITI period and trial onset, EtOH selectively altered striatal encoding of policy in neural populations recorded from males but not females ($p = 0.0016$ for the interaction between treatment, sex, and time). In EtOH male rats, the proportion of policy-encoding neurons decreased significantly faster than in their air-exposed controls (**Figure 5E**; $p = 0.00064$ for the interaction between treatment and time bin), whereas no significant differences in policy encoding were observed between EtOH and air-exposed females (**Figure 5H**).

Lastly, we examined how EtOH pre-exposure affects value updating during the outcome feedback period by regressing firing rates aligned with the onset of outcome presentation as linear functions of the action value of the chosen option (chosen value, Qc), while accounting for the effects of choice and outcome in the analysis (**Figures 5C, 5F and 5I**). The impact of EtOH on neural encoding of the chosen value significantly varied depending on sex ($p < 0.0001$ for the interaction among treatment, sex and time). We found a greater proportion of neurons encoded the chosen value at the beginning of outcome presentation (~0-400 ms) in EtOH male rats than their air controls (**Figure 5F**; $p = 0.017$ for the interaction between treatment and time). By contrast, the chosen value was represented in significantly fewer neurons in EtOH female rats than their air controls (**Figure 5I**; $p = 0.035$ for the main effect of treatment in female rats).

These results suggest that EtOH differentially alters the neural representation of state values and action policy during the distal action planning phase (i.e., ITI period and trial onset) and the neural representation of value updating of chosen actions during the outcome processing period in males and females.

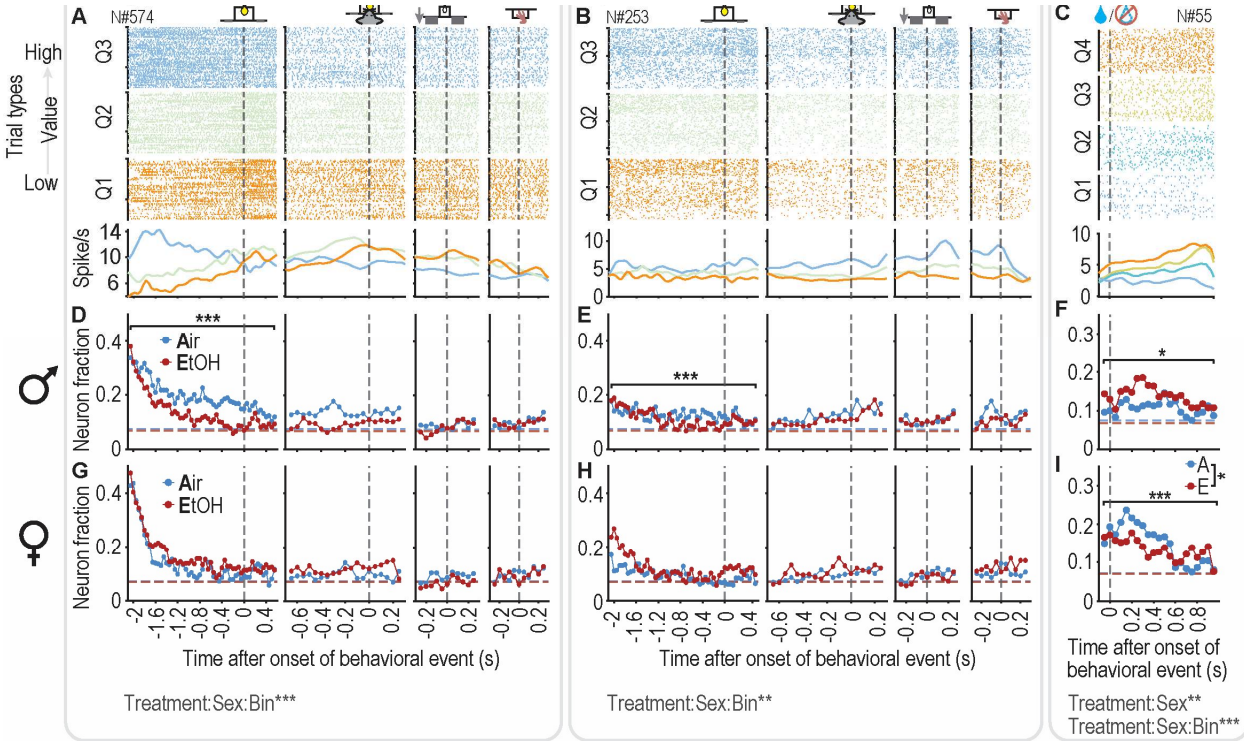

Figure 5. EtOH exposure alters striatal value encoding. (**A-C**) Example DMS unit from the air group shows spike activity correlated with value of state, ΣQ (A), and policy values, ΔQ (B), during action planning (i.e., inter-trial interval (ITI), trial onset, trial initiation, lever insertion) and action selection period, as well as chosen values (C) during outcome presentation. In the top raster plots in panels A and B, trials are grouped into three quantiles of the value (Q1, Q2, and Q3), with each quantile representing 33.3% of the data. In the top raster plots in panel C, trials are grouped into quartiles of the chosen action value (Q1, Q2, Q3, and Q4), with each quartile representing 25% of the data. At the bottom are spike density functions generated by applying a Gaussian kernel (δ = 100 ms) to the corresponding spike trains from A-C. (**D-F**) Fraction of neurons from male (♂) rats with activity significantly modulated by state value, ΣQ (D), and policy values, ΔQ (E), during action planning and action selection period, as well as chosen values (F) during outcome presentation. At D, ***$p < 0.0001$ interaction between treatment (EtOH v.s., Air) and time in trial onset epoch. At E, ***$p = 0.00064$ interaction between treatment and time in trial onset epoch. At F, *$p = 0.017$ interaction between treatment and time in outcome presentation. (**G-I**) Fraction of neurons from female (♀) rats with activity significantly modulated by state value, ΣQ (G), and policy values, ΔQ (H), during action planning and action selection period, as well as chosen values (I) during outcome presentation period. At I, *$p = 0.035$ main effect of treatment; ***$p = 0.00017$ interaction between treatment and time; *$p = 0.0016$ interaction between treatment and sex; ***$p < 0.0001$ interaction among treatment, sex, and time in outcome presentation. In the B-I, the horizontal dashed lines indicate the chance level (binomial test, $\alpha = 0.05$). Logistic regression with Generalized Linear Mixed-Effect Model for D-I. n = 241 neurons from 9 air male rats, 494 neurons from 8 EtOH male rats, 347 neurons from 5 air female rats, and 354 neurons from 5 EtOH female rats.

Discussion

In this study, we demonstrated that chronic EtOH vapor exposure during adulthood induces long-lasting behavioral deficits in rats, which manifest in a sex-specific manner and persist for over 10 weeks after withdrawal. Behavioral data from the dynaPRL, but not the simple PRL, task revealed that it is possible to predict sex and the history of EtOH exposure in male subjects based on their performance. In particular, cycles of high-dose EtOH exposure impaired reversal learning under conditions of high uncertainty and led to a general tendency to repeat the same choice regardless of outcome in male, but not female, rats. To elucidate the mechanisms underlying these behavioral deficits in males, we employed a reinforcement learning model. We found that chronic EtOH exposure increased both the outcome-mediated learning rate, which updates the value of chosen actions, and the forgetting rate after reward omission, which regulates the decay over time of the value of unchosen actions. Furthermore, EtOH exposure differentially altered the firing patterns of dorsomedial striatum (DMS) neurons in male and female rats. In males, EtOH exposure heightened the magnitude of outcome representations and enhanced the representation of chosen value, while it diminished responsiveness to choice, policy, and state signals, a pattern of changes that aligns with the observed reduced cognitive control of action-selection and deficits in outcome-specific value updating in EtOH males. Conversely, in females, EtOH exposure specifically enhanced neural encoding related to choice and prolonged the duration of outcome representations, without significantly impacting the encoding of other decision variables. Our findings shed light on the sex-dependent cognitive and neural computational mechanisms underlying drug-induced persistent deficits in cognitive flexibility.

Influences of EtOH exposure history on discrete aspects of learning

Human studies have linked chronic EtOH abuse and withdrawal with long-lasting cognitive impairments (39-41). However, in rodent studies, the evidence for enduring impacts of EtOH on cognitive flexibility has been mixed. Promising results have emerged from reversal tests within attention set-shifting tasks (13, 42) (but see also (43)) and the spatial Barnes maze (44), although, findings are less consistent in probabilistic reward reversal tasks (33, 34, 43, 45, 46). In our studies here, we also used two-choice probabilistic learning procedures, which rely on trial-and-error learning within sessions of hundreds of behavioral trials, and are thus amenable to computational modeling to characterize specific steps in learning processes (47), allowing for detection of the underlying cognitive mechanisms that may underlie EtOH-induced deficits, including some of those previously observed. Our findings from a standard probabilistic reversal learning (PRL) task showed no significant behavioral differences between pre and post EtOH exposure, consistent with prior studies in both rodents (33, 34) and humans (48). However, by varying uncertainty in reward probabilities in a new dynaPRL task (35, 36) and hence enhancing its complexity and challenge, we uncovered reliable EtOH-induced cognitive deficits that are sex-dependent. When facing sudden cognitive challenge in the dynaPRL task, EtOH males, but not females, exhibited a moderate but reliable deficit in reversal learning, i.e., a higher probability of repeating their choice, especially after reward omission in a high-uncertainty environment. Additionally, male EtOH rats exhibited less deviation from matching their choices to the inferred reward schedule (49). These findings suggest that EtOH rats, especially males, might favor exploitation over exploration, generally in line with human studies (50). Exploration-exploitation behaviors are modulated by

uncertainty and outcome novelty (51).

One plausible explanation for the EtOH-induced reduction in exploration is that chronic EtOH exposure either decreases uncertainty-mediated exploration (52) or promotes uncertainty avoidance following unrewarded choices, independent of learning (51). This may also explain why behavioral deficits with standard PRL, in which the uncertainty levels are low and constant, have been less frequently observed, whereas our dynaPRL might provide favorable conditions for observing deficits given the greater challenge to animals to adapt their learning process according to changing uncertainty. Indeed, humans with EtOH use disorder display aberrant decision-making amidst high ambiguity in the Iowa Gambling Task (53). The intolerance of uncertainty has also been associated with risky EtOH use behavior (54) and "hyperkatifeia" (55).

Another possible explanation for the reduced exploration behavior is that EtOH pre-exposure may impair the processing of outcomes, thereby diminishing outcome sensitivity and altering decision-making under uncertainty. To estimate outcome learning processes, we used reinforcement learning, a theoretical framework for describing the cognitive process of value-based decision-making (56) that has been informative for the dissection of pathological decision making in animal models of addiction (57). We showed that a history of chronic EtOH exposure in males enhanced the updating of chosen action value upon receiving rewards, indicating heightened reward sensitivity. It also facilitated the decay of the unchosen action value when unrewarded, reflecting a diminished emphasis on negative feedback on chosen action. These alternations of reward learning in males heightened the emphasis on selected actions regardless of their outcomes, leading to increased likelihood of behavioral exploitation. This result aligns well with RL modeling of behavior of alcohol dependent humans alcohol users (58). Importantly, we observed similar but dramatically weaker trends of changes in outcome processes in females, which further support the proposition that high behavioral perseveration and exploitation in males stem from hypersensitivity to distinct outcomes. Taken together, our results show that male rats are more susceptible to the negative impacts of chronic EtOH vapor exposure on uncertainty-driven behavioral flexibility than female rats within the dynaPRL task.

Human studies on sex differences in the impact of chronic EtOH on cognitive function have produced mixed results, with some showing greater, lesser, or indistinguishable effects in women compared to men (30, 59-61). The reasons underlying these variations in results are multifaceted and influenced by factors such as the type of cognitive task, age, socioeconomic status, hormonal differences, and drinking pattern (60, 61). Laboratory rodent models provide a setting in which many of these variables are either absent or controlled. In these studies, chronic intermittent ethanol (CIE) vapor exposure impairs goal-directed decision-making in male rats but not in females (62), in line with findings here. In related studies of the impact of prenatal alcohol, males also appear more susceptible to later cognitive deficits than females (63). The sex-specific effects of EtOH observed in the present work may be attributed to several factors. First, we identified sex-based differences in basal cognitive processes reflected in choice behavior and reaction times. This finding suggests potential basal sex differences in explore-exploit strategies and motivations, supporting the idea that male and females engaged different cognitive process to learn or perform the task (5, 64). These sex differences in behavior may modulate susceptibility to neuropsychological impairments (65, 66). Thus, our results suggest that a high exploitation trait may make males more vulnerable to the impact of EtOH on

cognitive flexibility. Additionally, inherent sex differences in processing uncertainty, especially under cognitive challenge and stress, could also contribute to these observed behavioral differences (67). We speculate that female rats might possess protective factors, such as sex-specific impacts on estrogen, which offer neuroprotective and anti-inflammatory benefits (68).

Sex-Specific neural adaptations to EtOH in the dorsomedial striatum

Numerous studies support a critical role for the DMS in encoding decision-related variables (1, 2, 4, 7-9, 23, 38) and controlling action selection in value-based decision tasks (4, 8, 9, 23, 25, 69, 70). The overall pattern of changes in neural encoding of these variables in male rats exposed to EtOH is congruent with the reduced exploration we observed. First, the rapid decay of neural choice signals in EtOH exposed rats during the outcome presentation phase might hamper the temporal credit assignment process (71-73), weakening response-outcome associations. Additionally, the EOH-induced enhancement of outcome encoding in males may lead to the asymmetric facilitation of chosen action value updating observed in our study (**Figure 3**), increasing the learning rate for chosen values and accelerating the decay of unchosen values without reward. As expected from the RL modeling results, we observed an enhanced representation of chosen value in the DMS after EtOH exposure, which could lead to an overemphasis on the chosen option over alternatives, resulting in greater behavioral perseveration, reduced exploration, and ultimately, a loss of cognitive flexibility. Finally, the analysis of state value encoding also is in-line with the behavioral observations. In the present study, state value indicates the richness of the environment, which is a strong indicator of block type (**Figure S10**). This information can facilitate the construction of a 'world model' to guide action selection or help reduce uncertainty in action selection. Therefore, the rapid decay of state value representation in EtOH-exposed animals during the choice preparatory period might impair the DMS's ability to integrate environmental information, potentially hindering flexible model-based action selection. This impairment of state value signals, alongside the disruption of choice signals, could serve as another parallel mechanism to exacerbate poor response-outcome associations, ultimately causing males to be less exploratory in decision-making.

Female rats, however, did not display this pattern of neural activity. Their basal neural responses to decision variables differed from males, and chronic EtOH exposure affected their neural encoding in a distinct manner. The dynamics of DMS encoding of choice and outcome during outcome presentation differed considerably in EtOH-exposed females compared to EtOH-naïve females. EtOH exposure in females led to a lower representation of chosen value signals compared to their air-treated controls, a pattern opposite to that observed in males. The observed sex differences in DMS sensitivity to EtOH are unlikely to result from basic physiological properties of the recorded cells, since these properties are similar across sexes (74). While estrous cycle can influence female neural physiology (74), our study's sampling across several weeks likely averages out these cycle-dependent changes. Despite detection of some EtOH-induced neural encoding differences, surprisingly, we observed no significant EtOH-induced behavioral differences in females, suggesting that female rats may possess compensatory mechanisms that mitigate the effects of EtOH exposure on neural encoding in the DMS, thereby preserving flexible value-based decision-making. Or, as mentioned above, female rat use of different behavioral strategies may make them less susceptible to EtOH-induced neural changes and EtOH may result in

different profiles of cognitive alterations than in males to which our task was not sensitive. Thus, more research with a broad assortment of cognitive tasks is needed to fully understand the sex-specific neural and behavioral adaptations to drug exposure in order to develop female-sensitive tasks that effectively capture these differences.

In summary, this study has unveiled sex-specific and enduring behavioral and neural consequences of chronic EtOH exposure, providing new insights into mechanisms underlying cognitive impairments associated with EtOH addiction. Chronic EtOH exposure persistently and differentially distorts neural computational processes in males and females. Female rats exhibited distinct basal behavioral and neural dynamics during decision making which suggest the engagement of different computational systems that may be differentially impacted by EtOH, thereby avoiding the cognitive inflexibility observed in male rats. These mechanisms identified in males may underlie some of the cognitive deficits observed in humans with alcohol use disorder. Note that low cognitive flexibility imposes a higher risk of alcohol use disorders (75). Importantly, behavioral and neural mechanisms for the changes we observed here after chronic EtOH might resemble those underlying loss of cognitive flexibility after chronic use of other drugs of abuse (76) or other psychiatric disorders (77).

Materials and Methods

Animals

17 Male and 10 female Long-Evans (LE) rats (10 wks old upon arrival, Envigo) were singly housed. Rats were kept in a temperature- and humidity-controlled environment with a light:dark cycle of 12:12 h (lights on at 7:00 a.m.). Experiments were conducted during their light cycle. All animal procedures were approved by the Johns Hopkins University Animal Care and Use Committee.

Chronic intermittent EtOH vapor exposure (CIE)

Rats were exposed to four cycles of EtOH vapor (n = 8 male and 5 female) or air (n = 9 male and 5 female) (14, 32). Each cycle consisted of 14 hrs of vapor exposure followed by a 10-hr withdrawal, repeated for 5 consecutive days. EtOH was volatilized by venting air through a heating flask containing 95% EtOH at a rate of 30-40 l/min. EtOH vapor was delivered to rats housed in Plexiglas chambers (La Jolla Alcohol Research, inc.). Blood was collected from the tip of tails in the middle and end of each round from all EtOH-exposed rats to assess blood EtOH concentrations (BECs) with an AM1 alcohol analyzer from Analox Inc. To control the stress of the blood collection, we pinched the tip of tails for air control rats instead of cutting the tissue and collecting blood. Our procedure produced a mean BEC of 224.95 ± 14.51 (S.E.M) in male and 257.2 ± 23.79 (S.E.M) mg/dl in female rats without significant sex differences (**Figure 1A**, inset), which is in line with previous findings (14, 32).

Probabilistic reversal learning (PRL) task

Rats undergoing behavioral testing were water-restricted to 90% of their *ad-libitum* weight. Behavioral experiments were performed in a soundproof customized modular operant chamber (Med Associates). The levers and a reward magazine were confined to the same wall of the chamber. Rats were initially trained to enter the magazine to collect reward by delivering a 100 µl reward (10% w/v sucrose in tap water) at random times with an average interval of 60 s (RT60). Subsequently, rats were trained to initiate a trial by entering the magazine, which triggered insertion of two levers into the chamber. Pressing a lever led to reward delivery (at this stage, both levers provided reward with 100% probability). Next, rats were trained on the reversal task, in which pressing one lever consistently delivered a reward (100% probability) and the other lever was not rewarded (0% probability). These action-outcome contingencies were reversed randomly 8 trials after the value of an exponential moving average over the past 8 choices (1 for correct and 0 for incorrect choice) crossed a 0.75 threshold. The block reversal probability was 10%. If the block was not switched within 20 trials of crossing the threshold, the computer forced the reversal. Thus, each rat had adequate trials to acquire current contingencies and could not predict the contingency reversal. One to three days after the deterministic reward schedule (100% and 0%) training, training in the full PRL task began. The reward probabilities for each lever were selected from 70% and 10%. Each trial in the PRL task began with illumination of the reward magazine. Upon a poke into the magazine (trial initiation), two levers flanking the left and right sides of the port were inserted after a variable delay (randomly selected between 100 ms to 1 s with 100 ms intervals). Rats were given 30-s to press a lever; failure to do so resulted in an invalid trial (<1% of trials for all rats) that was excluded from analysis. A lever press probabilistically resulted in either the CS- (0.5 s of white noise generated by MED Associated, ENV-225SM) with no reward delivery or the CS+ (two clicker sounds with a 0.1-s interval generated by MED Associated, ENV-135M) with ~33µl reward delivery in the magazine via a syringe pump (MED Associated, PHM-

100). Each trial ended either at the end of the CS- presentation or after completing the reward collection (exit from the first magazine entry after reward delivery). Trials were separated by a constant 2.5-s intertrial interval (ITI). Each behavioral session lasted ~ two hours. All rats experienced 15 sessions of PRL prior to the vapor procedure and on average ~10.07 sessions of PRL after the vapor procedure.

Dynamic probabilistic reversal learning (dynaPRL) task

Water-restricted rats performed this task in the same operant chamber as the original PRL task. The task structure was the same as PRL above except that reward probabilities assigned to each lever were drawn pseudorandomly from a set of blocks with paired probabilities [block1(neutral): (0.45, 0.45), block2(low contrast): (0.6, 0.3), block3(high contrast): (0.8, 0.1)] (**Figure 1F**). Note that each block type had a distinct probability contrast (block1, 1:1; block2, 2:1; block3, 8:1), while maintaining the same average reward probability (0.45), and that the outcome uncertainty within a block varied with the degree of probability contrast across the two levers, i.e., high uncertainty in a low contrast block, etc. The lever outcome contingencies were randomly reversed to one of the three blocks every 15-30 trials, except when a rat made four or more consecutive incorrect choices within certain trial blocks (block2: 60,30 or block3: 80,10); such trials were excluded from trial progress. When switching blocks, the transition rules imposed two restrictions: 1) the position of the lever with the highest reward probability should not be the same in two consecutive blocks even if the reward probability contrast changes, and 2) neutral blocks (block1) should not be repeated. Block transitions were categorized into three types based on the relative cognitive challenge level before and after each block transition. If the relative uncertainty was lower in the post-reversal block compared with the pre-reversal block, this was termed a 'Hi-Lo' transition (i.e., block2→block3, block1→block2, or block1→block3). If the relative uncertainty was higher in the post-reversal block compared with the pre-reversal block, this was termed a 'Lo-Hi' transition (i.e., block3→block2, or block2→block1, or block3→block1). If the relatively uncertainly was the same in the post-reversal block compared with the pre-reversal block while still shifting probabilities across the right and left levers, this was termed a 'Same' transition (i.e., block2→block2 or block3→block3). In addition, in this task, during non-neutral blocks (i.e., blocks 1 and 3) in 40% of the trials after the first post-reversal 12 trials, we probed the animal's response to unexpected changes in reward size by doubling or halving the reward size (66 or 16.5 μl) compared to the remaining 60% of the trials. However, EtOH exposure did not alter choice strategy relative to distinct reward sizes (**Figure S11**); thus, we do not focus on this feature in this report. Rats were trained on the final dynaPRL task for at least 10 sessions prior to surgery.

Electrode implantation

All rats were surgically implanted with custom-made drivable electrode arrays, with custom-designed 3D-printed pieces, sixteen insulated tungsten wires (50 um, A-M Systems), and two silver ground wires, soldered to two Plexon headstage connectors (Plexon Inc. TX) using standard stereotaxic procedures. Electrode arrays were placed unilaterally in the left or right DMS (AP: -0.1 mm; ML: 2.45 mm; DV: -4.7 mm from Bregma).

Recording and Spike sorting

Following 1 week of recovery, rats were trained on the dynaPRL task again until they became accustomed to performing while their headstage was connected via a patch cable to the commutator located in the center of the chamber ceiling. Once their completed trials/session reached at least 80% of their pre-surgery mean number of trials in a session (3-7 sessions), recording sessions began. Electrical signals and behavioral events were collected using Plexon

multichannel neural recording systems (MAP system;) or Plexon OmniPlex systems, with a 40 kHz sampling rate. Neural recordings at the same depth in the DMS were obtained for multiple sessions; if multiple sessions from the exact recording location were included in the following analysis, the data from the same channel was included only once. After that, electrodes were lowered by approximately 160 µm, and recording resumed in the new location at least one day later.

Individual units were isolated offline using principal component analysis in Offline Sorter (Plexon) as described previously (78). Auto-correlograms, cross-correlograms, the distribution of inter-spike intervals, L-ratio, and isolation distance were used to ensure good isolation of single units. Spike and behavioral event timestamps were exported to MATLAB (MathWorks) using customized scripts for further analysis.

Data Analysis and Statistics

Behavioral metrics

Reversal performance

Reversal performance was assessed by the probability of maintaining the same choice before the contingency change, denoted as *p(perseveration)*. *p(perseveration)* was computed for each trial during the last 5 trials before and the first 12 trials after the contingency change by the ratio of correct choices and the total number of blocks, except that if the pre-reversal block was a neutral block, in which neither side is correct, then a random side was selected as a pseudo-correct choice for the calculation. A single exponential decay function was fit to the p(preservation) for 1-12 trials after the contingency change,

$$p = d + a \times e\hat{}(-\lambda \times t) \quad (1)$$

where d, a, and λ represent the pre-reversal asymptotic value, the magnitude of change in *p(perseveration)* after the reversal, and the learning speed, respectively. The single exponential decay function was defined by MATLAB function *fittype* and curve fitting employed MATLAB function *fit* with lower boundary (0, 0.5, 0) and upper boundary (3, 2, 1) for parameters d, a, and λ correspondingly. Averaged parameter values across 9 iterations were used for each subject and session.

Matching performance, stay probabilities, and latencies

To assess how animals allocate their responses (left and right lever presses) to the local (within a block) reward probabilities of each choice option after the learning period (first 12 trials), we calculated the *matching score* (37) as a deviation of left response probability, *p(left),* from the reward probability given selection of the left choice, *p(reward|choice=left)*. Deviation < 0 is undermatching, and > 0 is overmatching.

We measured the overall stay probability by calculating the likelihood of repeating the same choice as the last trial. Conditional stay probabilities were calculated separately for the reward or unrewarded outcome in the last trial and denoted as *p(stay|win)* and *p(stay|lose)* correspondingly. The difference between win-stay and lose-shift (ΔWSLS) was obtained by the difference between *p(stay|win)* and *p(shift|lose)*.

The latencies (response times) measured included the time interval between stimulus (center light illumination or lever insertion) and response or reaction (magazine entry or lever press) to that

stimulus. As the distribution of latencies is highly skewed to the left (zero), a natural logarithm transformation was used to make the distributions of latencies more akin to normal distributions.

Classification analysis

Support vector machine (SVM)

An SVM was employed to decode group membership (CIE vs. Air), sex label (male vs. female), and assessment phase (only for PRL task, pre-vapor vs. post-vapor) from the behavior dataset (**Figure S2A**). SVM modeling was performed using the MATLAB function *fitclinear* for the classification of binary group membership and *fitecoc* for the classification of multi-subject labels. For both classification approaches, we used a total of 20 behavioral measures (columns) from each session consisting of (#1) reversal decay in post-reversal and (#2) matching score in pre-reversal, as well as 9 other behavioral measurements in pre- and post-block transitions, accordingly (#3 - #20 see table below).

Feature ID	Phase relative to block switch	Feature full name
#3	Prior to block switch (pre)	Stay probability
#4		Stay probability given win
#5		Stay probability given lose
#6		Difference between win-stay and lose-shift probabilities
#7		Logarithmic transformed choice latency
#8		Logarithmic transformed choice latency given stay
#9		Logarithmic transformed choice latency given switch
#10		Logarithmic transformed trial initiation latency given win
#11		Logarithmic transformed trial initiation latency given lose
#12	Post to block switch (post)	Stay probability
#13		Stay probability given win
#14		Stay probability given lose
#15		Difference between win-stay and lose-shift probabilities
#16		Logarithmic transformed choice latency
#17		Logarithmic transformed choice latency given stay
#18		Logarithmic transformed choice latency given switch
#19		Logarithmic transformed trial initiation latency given win
#20		Logarithmic transformed trial initiation latency given lose

The behavior dataset was randomly divided into two parts, a training set and a test set with a 90:10 ratio. Within the training set, a linear SVM model was optimized using 10-fold cross-validation. This approach involves dividing the training set into 10 equal parts, training the model on nine

parts, and validating it on the tenth part. This process is repeated ten times, with each part used once as the validation set. The results from these iterations are used to fine-tune the model parameters. The optimized SVM model was then evaluated on a test set to assess its generalizability and predictive performance. This train-test process was repeated 1000 times. True positive (TP), true negative (TN), false positive (FP, type I error), and false negative (FN, type II error) were calculated based on actual predicted data labels. For the multilabel classification, precision [TP/(TP+FP)] was used to assess model performance in predicting each class. To determine whether the observed decoding accuracy for group membership and sex label was significantly higher than what would be expected by chance, we conducted Monte Carlo significance tests, and the p-value was computed as follows:

$$p = \frac{NP + 1}{N + 1} \tag{2}$$

Here, *NP* is the number of iterations where the decoding precision of an SVM model trained with a shuffled dataset is higher than that of the observed dataset [precision (Shuffle) > precision (Data)] and *N* is the number of all iterations.

To assess the contribution of individual behavioral features to the classification of sex labels in EtOH-naïve control rats, we employed the feature selection method of maximum relevance and minimum redundancy (fMRMR) using MATLAB. We directly applied the fMRMR function to the behavioral dataset, which was previously used to train the SVM classifier. This method allowed us to identify the most relevant features for distinguishing between male and female labels while minimizing redundancy among them. The importance score for each feature was defined based on its relevance to the classification task and its redundancy with other features, providing insights into the key behavioral features contributing to the classification.

Uniform manifold approximation and projection (UMAP)

The manifold learning technique, UMAP, was employed to perform non-linear dimensional reduction of our behavior data matrix from above. We used the UMAP implementation provided in the MATLAB community *UMAP*.

Reinforcement learning models

Reinforcement learning (RL) models assume that choices stem from the cumulative outcomes of actions accrued over numerous trials. We fitted the choice and outcome data to eight distinct RL models: 1, standard Q-learning model; 2, Q-learning model with differential learning rates (Q-DFLr); 3, differential forgetting RL model (DF-RL); 4, DF-RL model with inverse temperature β (DF-RLwβ); 5, forgetting RL model (F-RL); 6, F-RL model with inverse temperature β (F-RLwβ); 7, win-stay lose-shift model (WSLS); 8, biased random choice model (RD).

For models 1 through 4, choices are assumed to follow learned values for each option, Q, according to softmax function, in which the probability of choice *k* on trial *t* is:

$$P_k(t) = \frac{e^{\beta Q_k(t)}}{\sum_{i=1}^{k} e^{\beta Q_k(t)}} \tag{3}$$

Here, β is the inverse temperature parameter that controls the level of stochasticity in the choice, with β = 0 resulting in equal probability over all choice options. β was set to a constant 1 for models 4 and 6.

The value updating for each model is described below:

Model 1: Q-learning model,

If $c(t)$ = L,

$$Q_L(t+1) = Q_L(t) + \alpha\big(\mathrm{r(t)} - Q_L(t)\big) \tag{4}$$
$$Q_R(t+1) = Q_R(t) \tag{5}$$

where $c(t)$ is the choice side (left, L or right, R) in the current trial, $r(t)$ is the outcome on that trial, and α is the learning rate, controls the update of the chosen action value according to the reward prediction error $r(t) - Q_L(t)$.

Model 2: Q-learning model with differential learning rates (Q-DFLr),

If $c(t)$ = L and $r(t)$ = 1,

$$Q_L(t+1) = Q_L(t) + \alpha^{+}\big(\mathrm{r(t)} - Q_L(t)\big) \tag{6}$$
$$Q_R(t+1) = (1 - \theta^{+})\, Q_R(t) \tag{7}$$

If $c(t)$ = L and $r(t)$ = 0,

$$Q_L(t+1) = Q_L(t) + \alpha^{-}\big(\mathrm{r(t)} - Q_L(t)\big) \tag{8}$$
$$Q_R(t+1) = (1 - \theta^{-})\, Q_R(t) \tag{9}$$

Here, value updates are governed by different learning rates for positive, α^{+}, or omitted, α^{-}, outcomes and the non-chosen option is decay by a forgetting rate, θ^{+} or θ^{-}, for rewarded and unrewarded trials respectively.

Model 3 and 4: Differential forgetting RL (DF-RL) and DF-RL with (model 4) or without (model 3) free β,

If $c(t)$ = L and $r(t)$ = 1,

$$Q_L(t+1) = \gamma_c Q_L(t) + k^{+} \tag{10}$$
$$Q_R(t+1) = \gamma_{uc} Q_R(t) \tag{11}$$

If $c(t)$ = L and $r(t)$ = 0,

$$Q_L(t+1) = \gamma_c Q_L(t) + k^{-} \tag{12}$$
$$Q_R(t+1) = \gamma_{uc} Q_R(t) \tag{13}$$

where κ^{+} and κ^{-} directly update the chosen action value in rewarded and unrewarded trials, respectively. γ_c and γ_{uc} are the forgetting factor in updating chosen and unchosen action values, respectively. Accordingly, forgetting rates θ_c and θ_{uc} in models 3 and 4 corresponds to $1-\gamma_c$ and $1-\gamma_{uc}$, respectively.

Model 5 and 6: Forgetting RL models (F-RL) and F-RL with (model 6) or without (model 5) free β,

If $c(t)$ = L and $r(t)$ = 1,

$$Q_L(t+1) = \gamma Q_L(t) + k^{+} \tag{14}$$

If $c(t)$ = L and $r(t)$ = 0,

$$Q_L(t+1) = \gamma Q_L(t) + k^- \tag{15}$$

For the unchosen right side,

$$Q_R(t+1) = \gamma Q_R(t) \tag{16}$$

where κ^+ and κ^- directly update the chosen action value in rewarded and unrewarded trials, respectively. γ is the forgetting factor in updating chosen and unchosen action values.

Models 7 and 8 directly describe the probability of each choice option on trial t, either subject to experienced outcomes (WSLS) or as a fixed proportion of choice.

Model 7: Win-stay lose-shift.

If $c(t)$ = L and $r(t)$ = 1,

$$P_L(t+1) = 1 - \epsilon/2 \tag{17}$$

$$P_R(t+1) = \epsilon/2 \tag{18}$$

If $c(t)$ = L and $r(t)$ = 0,

$$P_L(t+1) = \epsilon/2 \tag{19}$$

$$P_R(t+1) = 1 - \epsilon/2 \tag{20}$$

Where ϵ controls the randomness of the stay probability.

Model 8: Biased random choice.

$$P_L(t+1) = \varphi \tag{21}$$

$$P_R(t+1) = 1 - \varphi \tag{22}$$

Where φ controls the bias of choice probability.

Hierarchical model fitting and model comparisons

We used the Matlab Stan interface (https://mc-stan.org/users/interfaces/matlab-stan) to construct and sample from hierarchical RL models, in which group-level hyperparameters govern individual rat parameters, to avoid overfitting to inter-rat noise. For a given model, we used Stan to draw consecutive samples from the posterior density of the model parameters using Hamiltonian Markov chain Monte Carlo (HMC), giving density estimates for the best fitting model parameters for the aggregate choice and outcome data across all sessions and rats. For each model, rat-level parameters were drawn from a group-level distribution with group mean, μ, group mean difference, δ, and variance σ. To quantify the evidence of group differences, the Directed Bayes Factor (dBF) was employed. This measure calculates the ratio of the proportion of the distribution of δ (group difference) that is above zero to the proportion that is below zero as in (79). Given the difficulties in sampling from hierarchical RL models, parameters were transformed such that sample were drawn in an unconstrained space and later transformed into bounded values using the fast approximation of the cumulative normal distribution function, *phi_approx*. We set normally distributed priors for all parameters at the rat level: N(0, 1) for all parameters, except β, which was set at N(0, 10), while group-level hyperparameters were given a prior normal prior distribution of N(0, 2) for μ and δ; and a gamma distribution of (2, 0.5) for σ. Hierarchical models for each of the

candidate RL models were sampled for 2,000 iterations after 1,000 warm-up draws for four chains in parallel. All other settings adhered to the default configuration.

Predictive accuracy of each model was estimated using leave-one-out cross validation with Pareto-smoothed importance sampling (psis-loo) using the model log likelihood computed for each of the sampled parameters on the actual choice and outcome data. For simulations of behavior, we took the mean of the posterior samples (i.e. over the 2000 post-warm up draws) for each rat-level model parameter as the best-fitting estimate for that individual subject. Using these best-fitting model parameters, we simulated choice behavior on the actual task structure. A reversal analysis, analogous to the one used on the empirical data, was applied to these simulated data to probe whether the model could recapitulate the key differences in choice behavior observed in the empirical data.

Multiple linear regression

To examine how neural spike activity was modulated by different behavioral and decision variables, we employed multiple linear regression models to fit spike rates of each unit as a linear function of various factors, such as rat's choice, reward outcome, and estimated action values as in (2, 38). All trials in each session other than those with choice omission were included in the analysis. The number of trials was matched across groups (**Figure S7D**). To investigate the evolution of neural spike modulation within a temporal window surrounding key behavioral events, including lever insertion, action selection, and outcome presentation, we aligned the spikes of each neuron to the onset of these behavioral events. We then quantified the firing rates of each neuron using a sliding time bin approach. Specifically, we employed a 100-ms time bin with 50-ms steps or a 200-ms time bin with 100-ms steps. This analysis was conducted within predefined temporal windows relative to trial onset, trial initiation, lever insertion, lever press, and outcome, presentation, which were set at -2300 ms to 600 ms, -700 ms to 300 ms, -500 ms to 500 ms, -500 ms to 500 ms, and -100 ms to 1000 ms, respectively. To control for the temporal auto-correlation among spikes, we added auto-regression (AR) terms as in Shin et al. (38) for all of the following regression models. To investigate the influence of both current and preceding choices, outcomes, and their interaction on the spike activity of individual neurons, we employed a regression model to delineate the firing rate (S) as a function of choice direction (C; ipsi or contra-lateral to implant side), outcome (O; rewarded or unrewarded), and their interaction (I) in current trial t and last trial t-1, along with autoregressive components (AR), as the following

$$S(t) = \beta_0 + \beta_1 \times C(t) + \beta_2 \times C(t-1) + \beta_3 \times O(t) + \beta_4 \times O(t-1) + \beta_5 \times I(t) + \beta_6 \times I(t-1) + AR + \varepsilon(t) \quad (23)$$

where β_0 through β_6 represent standardized regression coefficients, and the error term is $\epsilon(t)$. AR term was defined as follows:

$$AR = \beta_7 \times S(t-1) + \beta_8 \times S(t-2) + \beta_9 \times S(t-3) \quad (24)$$

To examine how chosen value-related signals are represented by neurons at the time that choice and outcome are revealed, we used the following regression:

$$S(t) = \beta_0 + \beta_1 \times C(t) + \beta_2 \times O(t) + \beta_3 \times I(t) + \beta_4 \times Q_c(t) + AR + \varepsilon(t) \quad (25)$$

where Q_c is the chosen value estimated by the best fit RL model.

To investigate how different types of value signals are encoded in neuronal activity, we used a regression model as follows:

$$S(t) = \beta_0 + \beta_1 \times C(t) + \beta_2 \times \Sigma Q(t) + \beta_3 \times \Delta Q(t) + \beta_4 \times Q_c(t) + AR + \varepsilon(t) \quad (26)$$

Where ΣQ is the summation of action values of ipsilateral (Q_{ipsi}) and contralateral choice (Q_{contra}) and ΔQ is the difference between Q_{ipsi} and Q_{contra}, which can approximate state value and policy, respectively (2, 38).

Generalized linear mixed-effect models (GLMMs)

To understand the impact of EtOH treatment, sex and time during the epoch on the proportion of neural responses to specific decision variables described above, we employed a Generalized Linear Mixed-Effects Model (GLMM) to account for the random effects associated with individual animals and specific neurons within those animals. The model is described by the following Wikison notation:

$$RS \sim Group * sex * bin + (1|RatID) + (1|RatID{:}NeuronID) \quad (27)$$

Where *RS* denotes significance of response, a binary variable indicating whether the response is modulated significantly, as determined by the preceding multiple linear regression. The fixed effects in the model include *Group* (representing the experimental conditions, such as EtOH or Air), *bin* (representing time bins of spike activity), and the interaction term *Group:bin* (representing the interaction between the experimental conditions and time bins). *RatID* and *NeuronID* are random effects. The term "*1|RatID*" introduces a random intercept for each animal, allowing for the acknowledgment of variability in baseline response levels across animals. The term "*1|RatID:NeuronID*" adds another layer of random intercepts, this time to account for variability at the neuron level within each animal, acknowledging that responses may differ not just between animals but also among neurons within the same animal. If a sex-dependent interaction was observed, we conducted additional GLMM analyses within each sex as a post-hoc test, using the following Wilkison notation

$$RS \sim Group * bin + (1|RatID) + (1|RatID{:}NeuronID) \quad (27)$$

To fit this model, we utilize the *fitglme* function in MATLAB with a 'Binomial' distribution and a 'Logit' link function.

Other statistical analysis

Behavioral data with normal distribution or relatively small sample size were analyzed using paired t-tests, one-way ANOVA with repeated measures (one-way RM ANOVA), or mixed effect ANOVA (also known as two-way ANOVA with repeated measures, two-way RM ANOVA) followed by Bonferroni post hoc test. Non-parametric Wilcoxon tests were employed when indicated by non-normal distribution of the behavioral data. Despite potential deviations from normality, a mixed-effect ANOVA (considering both within-sample and between-sample effects) was applied to some behavioral data, given the method's robustness against such violations. The comparison of electrophysiological data encoding distinct properties, characterized by a hierarchical structure with varying numbers of neurons per animal, were analyzed using Generalized Linear Mixed Models (GLMMs) to account for variability at the animal and neuron levels. Additionally, to determine the chance level for the proportion of neurons significantly modulated by decision variables, a binomial test with $\alpha = 0.05$ was conducted. All conducted in MATLAB. A p-value < 0.05 was used as the criterion to determine statistical significance unless

noted otherwise. Data analyzed using parametric tests are expressed as the mean ± standard error of the mean (SEM). For data analyzed with non-parametric tests, we present the median, with the interquartile range (IQR) represented by the 1st (25th percentile) and 3rd quartiles (75th percentile). The whiskers on the box plot indicate the range of the data within 1.5 times the IQR from the quartiles, providing a clear visualization of the data distribution and variability.

References

1. M. Ito, K. Doya, Validation of decision-making models and analysis of decision variables in the rat basal ganglia. *J Neurosci* **29**, 9861-9874 (2009).
2. H. Kim, J. H. Sul, N. Huh, D. Lee, M. W. Jung, Role of striatum in updating values of chosen actions. *J Neurosci* **29**, 14701-14712 (2009).
3. B. Lau, P. W. Glimcher, Value representations in the primate striatum during matching behavior. *Neuron* **58**, 451-463 (2008).
4. A. Y. Wang, K. Miura, N. Uchida, The dorsomedial striatum encodes net expected return, critical for energizing performance vigor. *Nat Neurosci* **16**, 639-647 (2013).
5. J. Cox, A. R. Minerva, W. T. Fleming, C. A. Zimmerman, C. Hayes, S. Zorowitz, A. Bandi, S. Ornelas, B. McMannon, N. F. Parker, I. B. Witten, A neural substrate of sex-dependent modulation of motivation. *Nature Neuroscience* **26**, 274-284 (2023).
6. A. C. Burton, G. B. Bissonette, N. T. Lichtenberg, V. Kashtelyan, M. R. Roesch, Ventral Striatum Lesions Enhance Stimulus and Response Encoding in Dorsal Striatum. *Biological psychiatry* **75**, 132-139 (2014).
7. E. Y. Kimchi, M. Laubach, Dynamic Encoding of Action Selection by the Medial Striatum. *The Journal of Neuroscience* **29**, 3148-3159 (2009).
8. S. Nonomura, K. Nishizawa, Y. Sakai, Y. Kawaguchi, S. Kato, M. Uchigashima, M. Watanabe, K. Yamanaka, K. Enomoto, S. Chiken, H. Sano, S. Soma, J. Yoshida, K. Samejima, M. Ogawa, K. Kobayashi, A. Nambu, Y. Isomura, M. Kimura, Monitoring and Updating of Action Selection for Goal-Directed Behavior through the Striatal Direct and Indirect Pathways. *Neuron* **99**, 1302-1314 e1305 (2018).
9. D. Lee, H. Seo, M. W. Jung, Neural basis of reinforcement learning and decision making. *Annu Rev Neurosci* **35**, 287-308 (2012).
10. B. J. Mason, Emerging pharmacotherapies for alcohol use disorder. *Neuropharmacology* **122**, 244-253 (2017).
11. V. Manning, S. Wanigaratne, D. Best, R. G. Hill, L. J. Reed, D. Ball, J. Marshall, M. Gossop, J. Strang, Changes in Neuropsychological Functioning during Alcohol Detoxification. *European Addiction Research* **14**, 226-233 (2008).
12. S. Loeber, T. Duka, H. Welzel Márquez, H. Nakovics, A. Heinz, K. Mann, H. Flor, Effects of Repeated Withdrawal from Alcohol on Recovery of Cognitive Impairment under Abstinence and Rate of Relapse. *Alcohol and Alcoholism* **45**, 541-547 (2010).
13. M. W. Meinhardt, S. Pfarr, G. Fouquet, C. Rohleder, M. L. Meinhardt, J. Barroso-Flores, R. Hoffmann, J. Jeanblanc, E. Paul, K. Wagner, A. C. Hansson, G. Köhr, N. Meier, O. von Bohlen und Halbach, R. L. Bell, H. Endepols, B. Neumaier, K. Schönig, D. Bartsch, M. Naassila, R. Spanagel, W. H. Sommer, Psilocybin targets a common molecular mechanism for cognitive impairment and increased craving in alcoholism. *Science Advances* **7**, eabh2399 (2021).
14. R. Renteria, E. T. Baltz, C. M. Gremel, Chronic alcohol exposure disrupts top-down control over basal ganglia action selection to produce habits. *Nat Commun* **9**, 211 (2018).
15. L. H. Corbit, H. Nie, P. H. Janak, Habitual alcohol seeking: time course and the contribution of subregions of the dorsal striatum. *Biol Psychiatry* **72**, 389-395 (2012).
16. G. Dominguez, C. Belzung, C. Pierard, V. David, N. Henkous, L. Decorte, N. Mons, D. Beracochea, Alcohol withdrawal induces long-lasting spatial working memory impairments: relationship with changes in corticosterone response in the prefrontal cortex. *Addiction Biology* **22**, 898-910 (2017).

17. Y. Ehinger, Z. Zhang, K. Phamluong, D. Soneja, K. M. Shokat, D. Ron, Brain-specific inhibition of mTORC1 eliminates side effects resulting from mTORC1 blockade in the periphery and reduces alcohol intake in mice. *Nature Communications* **12**, 4407 (2021).
18. Y. Cheng, C. C. Y. Huang, T. Ma, X. Wei, X. Wang, J. Lu, J. Wang, Distinct synaptic strengthening of the striatal direct and indirect pathways drives alcohol consumption. *Biol Psychiatry* **81**, 918-929 (2017).
19. Y. Cheng, X. Xie, J. Lu, H. Gangal, W. Wang, S. Melo, X. Wang, J. Jerger, K. Woodson, E. Garr, Y. Huang, P. Janak, J. Wang, Optogenetic induction of orbitostriatal long-term potentiation in the dorsomedial striatum elicits a persistent reduction of alcohol-seeking behavior in rats. *Neuropharmacology* **191**, 108560 (2021).
20. T. Ma, Y. Cheng, E. Roltsch Hellard, X. Wang, J. Lu, X. Gao, C. C. Y. Huang, X. Y. Wei, J. Y. Ji, J. Wang, Bidirectional and long-lasting control of alcohol-seeking behavior by corticostriatal LTP and LTD. *Nat Neurosci* **21**, 373-383 (2018).
21. B. Bloem, R. Huda, K.-I. Amemori, A. S. Abate, G. Krishna, A. L. Wilson, C. W. Carter, M. Sur, A. M. Graybiel, Multiplexed action-outcome representation by striatal striosome-matrix compartments detected with a mouse cost-benefit foraging task. *Nature Communications* **13**, (2022).
22. F. Gore, M. Hernandez, C. Ramakrishnan, A. K. Crow, R. C. Malenka, K. Deisseroth, Orbitofrontal cortex control of striatum leads economic decision-making. *Nature Neuroscience*, (2023).
23. K. Samejima, Y. Ueda, K. Doya, M. Kimura, Representation of action-specific reward values in the striatum. *Science* **310**, 1337-1340 (2005).
24. L. H. Corbit, P. H. Janak, Posterior dorsomedial striatum is critical for both selective instrumental and Pavlovian reward learning. *The European journal of neuroscience* **31**, 1312-1321 (2010).
25. S. S. Bolkan, I. R. Stone, L. Pinto, Z. C. Ashwood, J. M. Iravedra Garcia, A. L. Herman, P. Singh, A. Bandi, J. Cox, C. A. Zimmerman, J. R. Cho, B. Engelhard, J. W. Pillow, I. B. Witten, Opponent control of behavior by dorsomedial striatal pathways depends on task demands and internal state. *Nature Neuroscience* **25**, 345-357 (2022).
26. H. H. Yin, S. B. Ostlund, B. J. Knowlton, B. W. Balleine, The role of the dorsomedial striatum in instrumental conditioning. *The European journal of neuroscience* **22**, 513-523 (2005).
27. R. M. Shansky, Behavioral neuroscience's inevitable SABV growing pains. *Trends in Neurosciences*, (2024).
28. A. K. Radke, E. A. Sneddon, R. M. Frasier, F. W. Hopf, Recent Perspectives on Sex Differences in Compulsion-Like and Binge Alcohol Drinking. *International Journal of Molecular Sciences* **22**, 3788 (2021).
29. A. Flores-Bonilla, H. N. Richardson, Sex Differences in the Neurobiology of Alcohol Use Disorder. *Alcohol Res* **40**, 04 (2020).
30. N. L. Zabik, J. U. Blackford, Sex and sobriety: Human brain structure and function in AUD abstinence. *Alcohol*, (2024).
31. R. van den Bos, J. Homberg, L. de Visser, A critical review of sex differences in decision-making tasks: Focus on the Iowa Gambling Task. *Behavioural Brain Research* **238**, 95-108 (2013).

32. N. W. Gilpin, A. D. Smith, M. Cole, F. Weiss, G. F. Koob, H. N. Richardson, Operant behavior and alcohol levels in blood and brain of alcohol-dependent rats. *Alcohol Clin Exp Res* **33**, 2113-2123 (2009).
33. C. L. Pickens, M. Gallo, H. Fisher, A. Pajser, M. H. Ray, Alcohol Consumption during Adulthood Does Not Impair Later Go/No-Go Reversal Learning in Male Rats. *NeuroSci* **2**, 166-176 (2021).
34. C. G. Aguirre, A. Stolyarova, K. Das, S. Kolli, V. Marty, L. Ray, I. Spigelman, A. Izquierdo, Sex-dependent effects of chronic intermittent voluntary alcohol consumption on attentional, not motivational, measures during probabilistic learning and reversal. *PLOS ONE* **15**, e0234729 (2020).
35. C. D. Grossman, B. A. Bari, J. Y. Cohen, Serotonin neurons modulate learning rate through uncertainty. *Current Biology* **32**, 586-599 e587 (2021).
36. A. Funamizu, M. Ito, K. Doya, R. Kanzaki, H. Takahashi, Uncertainty in action - value estimation affects both action choice and learning rate of the choice behaviors of rats. *European Journal of Neuroscience* **35**, 1180-1189 (2012).
37. E. Trepka, M. Spitmaan, B. A. Bari, V. D. Costa, J. Y. Cohen, A. Soltani, Entropy-based metrics for predicting choice behavior based on local response to reward. *Nature Communications* **12**, 6567 (2021).
38. E. J. Shin, Y. Jang, S. Kim, H. Kim, X. Cai, H. Lee, J. H. Sul, S.-H. Lee, Y. Chung, D. Lee, M. W. Jung, Robust and distributed neural representation of action values. *eLife* **10**, (2021).
39. S. Chanraud, M. Reynaud, M. Wessa, J. Penttilä, N. Kostogianni, A. Cachia, E. Artiges, F. Delain, M. Perrin, H.-J. Aubin, Y. Cointepas, C. Martelli, J.-L. Martinot, Diffusion Tensor Tractography in Mesencephalic Bundles: Relation to Mental Flexibility in Detoxified Alcohol-Dependent Subjects. *Neuropsychopharmacology* **34**, 1223-1232 (2009).
40. L. D. Vanes, R. J. Holst, J. M. Jansen, W. Brink, J. Oosterlaan, A. E. Goudriaan, Contingency Learning in Alcohol Dependence and Pathological Gambling: Learning and Unlearning Reward Contingencies. *Alcoholism: Clinical and Experimental Research* **38**, 1602-1610 (2014).
41. D. Jokisch, P. Roser, G. Juckel, I. Daum, C. Bellebaum, Impairments in learning by monetary rewards and alcohol-associated rewards in detoxified alcoholic patients. *Alcohol Clin Exp Res* **38**, 1947-1954 (2014).
42. S. Kroener, P. J. Mulholland, N. N. New, J. T. Gass, H. C. Becker, L. J. Chandler, Chronic Alcohol Exposure Alters Behavioral and Synaptic Plasticity of the Rodent Prefrontal Cortex. *PLoS ONE* **7**, e37541 (2012).
43. K. A. Badanich, H. C. Becker, J. J. Woodward, Effects of chronic intermittent ethanol exposure on orbitofrontal and medial prefrontal cortex-dependent behaviors in mice. *Behav Neurosci* **125**, 879-891 (2011).
44. A. Kuzmin, S. Liljequist, J. Meis, V. Chefer, T. Shippenberg, G. Bakalkin, Repeated moderate-dose ethanol bouts impair cognitive function in Wistar rats. *Addiction Biology* **17**, 132-140 (2012).
45. C. A. Dannenhoffer, M. M. Robertson, V. A. Macht, S. M. Mooney, C. A. Boettiger, D. L. Robinson, Chronic alcohol exposure during critical developmental periods differentially impacts persistence of deficits in cognitive flexibility and related circuitry. *International Review of Neurobiology* **160**, 117-173 (2021).

46. M. H. Ray, T. Hite, M. Gallo, C. L. Pickens, Operant over-responding is more sensitive than reversal learning for revealing behavioral changes after withdrawal from alcohol consumption. *Physiol Behav* **196**, 176-184 (2018).
47. A. Izquierdo, J. L. Brigman, A. K. Radke, P. H. Rudebeck, A. Holmes, The neural basis of reversal learning: An updated perspective. *Neuroscience* **345**, 12-26 (2017).
48. B. Bağci, S. Düsmez, N. Zorlu, G. Bahtiyar, S. Isikli, A. Bayrakci, A. Heinz, D. J. Schad, M. Sebold, Computational analysis of probabilistic reversal learning deficits in male subjects with alcohol use disorder. *Frontiers in Psychiatry* **13**, (2022).
49. K. Iigaya, Y. Ahmadian, L. P. Sugrue, G. S. Corrado, Y. Loewenstein, W. T. Newsome, S. Fusi, Deviation from the matching law reflects an optimal strategy involving learning over multiple timescales. *Nature Communications* **10**, (2019).
50. L. S. Morris, K. Baek, P. Kundu, N. A. Harrison, M. J. Frank, V. Voon, Biases in the Explore–Exploit Tradeoff in Addictions: The Role of Avoidance of Uncertainty. *Neuropsychopharmacology* **41**, 940-948 (2016).
51. J. Cockburn, V. Man, W. A. Cunningham, J. P. O'Doherty, Novelty and uncertainty regulate the balance between exploration and exploitation through distinct mechanisms in the human brain. *Neuron* **110**, 2691-2702.e2698 (2022).
52. S. J. Gershman, Uncertainty and Exploration. *Decision (Wash D C)* **6**, 277-286 (2019).
53. A. E. Goudriaan, J. Oosterlaan, E. de Beurs, W. van den Brink, Decision making in pathological gambling: A comparison between pathological gamblers, alcohol dependents, persons with Tourette syndrome, and normal controls. *Cognitive Brain Research* **23**, 137-151 (2005).
54. M. E. Oglesby, B. J. Albanese, J. Chavarria, N. B. Schmidt, Intolerance of Uncertainty in Relation to Motives for Alcohol Use. *Cognitive Therapy and Research* **39**, 356-365 (2015).
55. G. F. Koob, P. Powell, A. White, Addiction as a Coping Response: Hyperkatifeia, Deaths of Despair, and COVID-19. *American Journal of Psychiatry* **177**, 1031-1037 (2020).
56. N. D. Daw, Y. Niv, P. Dayan, "Actions, policies, values and the basal ganglia" (2005).
57. S. M. Groman, S. L. Thompson, D. Lee, J. R. Taylor, Reinforcement learning detuned in addiction: integrative and translational approaches. *Trends in Neurosciences* **45**, 96-105 (2022).
58. S. B. Beylergil, A. Beck, L. Deserno, R. C. Lorenz, M. A. Rapp, F. Schlagenhauf, A. Heinz, K. Obermayer, Dorsolateral prefrontal cortex contributes to the impaired behavioral adaptation in alcohol dependence. *Neuroimage Clin* **15**, 80-94 (2017).
59. S. J. Nixon, R. Prather, B. Lewis, "Chapter 16 - Sex differences in alcohol-related neurobehavioral consequences" in *Handbook of Clinical Neurology,* E. V. Sullivan, A. Pfefferbaum, Eds. (Elsevier, 2014), vol. 125, pp. 253-272.
60. R. Fama, A. P. Le Berre, E. V. Sullivan, Alcohol's Unique Effects on Cognition in Women: A 2020 (Re)view to Envision Future Research and Treatment. *Alcohol Res* **40**, 03 (2020).
61. J. F. Van den Berg, B. Dogge, N. Kist, R. M. Kok, K. Van der Hiele, Gender Differences in Cognitive Functioning in Older Alcohol-Dependent Patients. *Substance Use & Misuse* **52**, 574-580 (2017).
62. J. M. Barker, K. G. Bryant, J. I. Osborne, L. J. Chandler, Age and Sex Interact to Mediate the Effects of Intermittent, High-Dose Ethanol Exposure on Behavioral Flexibility. *Frontiers in Pharmacology* **8**, 450 (2017).

63. J. Chandrasekaran, B. Jacquez, J. Wilson, J. L. Brigman, Reinforcer value moderates the effects of prenatal alcohol exposure on learning and reversal. *Frontiers in Neuroscience* **17**, (2023).
64. C. S. Chen, E. Knep, A. Han, R. B. Ebitz, N. M. Grissom, Sex differences in learning from exploration. *eLife* **10**, (2021).
65. J. D. Salamone, I. Koychev, M. Correa, P. McGuire, Neurobiological basis of motivational deficits in psychopathology. *European Neuropsychopharmacology* **25**, 1225-1238 (2015).
66. John D. Salamone, M. Correa, The Mysterious Motivational Functions of Mesolimbic Dopamine. *Neuron* **76**, 470-485 (2012).
67. H. Lei, Y. Mochizuki, C. Chen, K. Hagiwara, M. Hirotsu, T. Matsubara, S. Nakagawa, Sex difference in the weighting of expected uncertainty under chronic stress. *Scientific Reports* **11**, 8700 (2021).
68. M. E. Jung, M. B. Gatch, J. W. Simpkins, Estrogen Neuroprotection Against the Neurotoxic Effects of Ethanol Withdrawal: Potential Mechanisms. *Experimental Biology and Medicine* **230**, 8-22 (2005).
69. L. H. Tai, A. M. Lee, N. Benavidez, A. Bonci, L. Wilbrecht, Transient stimulation of distinct subpopulations of striatal neurons mimics changes in action value. *Nat Neurosci* **15**, 1281-1289 (2012).
70. J. Lee, B. L. Sabatini, Striatal indirect pathway mediates exploration via collicular competition. *Nature* **599**, 645-649 (2021).
71. H. Kim, D. Lee, M. W. Jung, Signals for previous goal choice persist in the dorsomedial, but not dorsolateral striatum of rats. *J Neurosci* **33**, 52-63 (2013).
72. S. Yagishita, A. Hayashi-Takagi, G. C. Ellis-Davies, H. Urakubo, S. Ishii, H. Kasai, A critical time window for dopamine actions on the structural plasticity of dendritic spines. *Science* **345**, 1616-1620 (2014).
73. E. S. Her, N. Huh, J. Kim, M. W. Jung, Neuronal activity in dorsomedial and dorsolateral striatum under the requirement for temporal credit assignment. *Sci Rep* **6**, 27056 (2016).
74. J. A. Willett, J. Cao, A. Johnson, O. H. Patel, D. M. Dorris, J. Meitzen, The estrous cycle modulates rat caudate–putamen medium spiny neuron physiology. *European Journal of Neuroscience* **52**, 2737-2755 (2020).
75. T. A. Shnitko, S. W. Gonzales, N. Newman, K. A. Grant, Behavioral Flexibility in Alcohol‐Drinking Monkeys: The Morning After. *Alcoholism: Clinical and Experimental Research* **44**, 729-737 (2020).
76. S. M. Groman, K. M. Rich, N. J. Smith, D. Lee, J. R. Taylor, Chronic exposure to methamphetamine disrupts reinforcement-based decision making in rats. *Neuropsychopharmacology* **43**, 770-780 (2018).
77. V. M. Brown, L. Zhu, A. Solway, J. M. Wang, K. L. McCurry, B. King-Casas, P. H. Chiu, Reinforcement Learning Disruptions in Individuals With Depression and Sensitivity to Symptom Change Following Cognitive Behavioral Therapy. *JAMA Psychiatry* **78**, 1113-1122 (2021).
78. Y. Vandaele, N. R. Mahajan, D. J. Ottenheimer, J. M. Richard, S. P. Mysore, P. H. Janak, Distinct recruitment of dorsomedial and dorsolateral striatum erodes with extended training. *Elife* **8**, (2019).
79. A. Wiehler, K. Chakroun, J. Peters, Attenuated Directed Exploration during Reinforcement Learning in Gambling Disorder. *The Journal of Neuroscience* **41**, 2512-2522 (2021).

Acknowledgements

We thank members of the Janak lab for helpful discussions on experimental findings, and Nikhila Kalapatapu, Tina Feng, and Jeniffer Min for assisting on a subset of behavioral experiments.

Funding:

National Institutes of Health grant R01AA026306 (PHJ)
National Institutes of Health grant R01AA031609 (PHJ)
National Institutes of Health grant R01DA035943 (PHJ)
JHU Kavli NDI Fellowship (YC)
Intramural Research Program of the National Institute of Mental Health ZIAMH002983 (AJL)

Author contributions:
Conceptualization: YC and PHJ
Methodology: YC, AL, DL, and PHJ
Investigation: YC and RM
Visualization: YC
Supervision: PHJ
Writing: YC, RM, AL, DL and PHJ

Competing interests: The authors declare that they have no competing interests.

Data and materials availability: The datasets generated for this manuscript are available on request to ycheng62@jhu.edu and patricia.janak@jhu.edu. The code used in this manuscript is availble on github upon publication.

Supplementary Materials for

Chronic Ethanol Exposure Produces Sex-Dependent Impairments in Value Computations in the Striatum

Cheng *et al*.

*Corresponding author. Email: patricia.janak@jhu.edu, ycheng62@jhu.edu

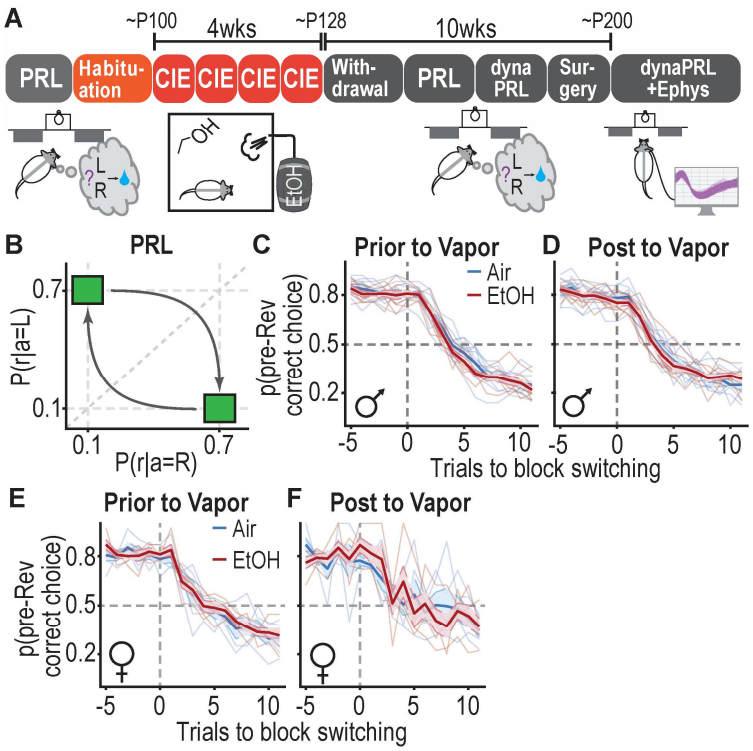

Figure S1. Protracted withdrawal from chronic ethanol exposure does not impair reversal learning in traditional probabilistic reversal learning task. (**A**) Experimental timeline: Rats underwent training in a simple probabilistic reversal learning (PRL) task to counterbalance their behaviors prior to 4-week chronic ethanol vapor exposure (CIE) procedure with a preceding habituation phase for acclimatization to ethanol vapor and experimental environment. Following 10-15 days of withdrawal from CIE, rats were retrained with the PRL task and then introduced to a new task, the dynamic probabilistic reversal learning (dynaPRL) task. 16-channel drivable electrodes were implanted in dorsomedial striatum of rats. After recovery and retraining, single-unit activity was recorded. (**B**) Choice reward probabilities for the PRL task. The reward probabilities associated with each lever switched drawing probability from a set (0.7, 0.1). (**C,D**) Probability of choosing pre-reversal preferred lever before (C) and after (D) CIE vapor in males. N = 9 male air rats, 8 male CIE rats. (**E**,**F**) Probability of choosing pre-reversal preferred lever before (E) and after (F) CIE vapor in females. N = 5 female Air rats, 5 female CIE rats. Bold lines represent the group average and faded lines represent individual rats. Vertical dashed line indicates the last trial before block switching (t=0) for D-G.

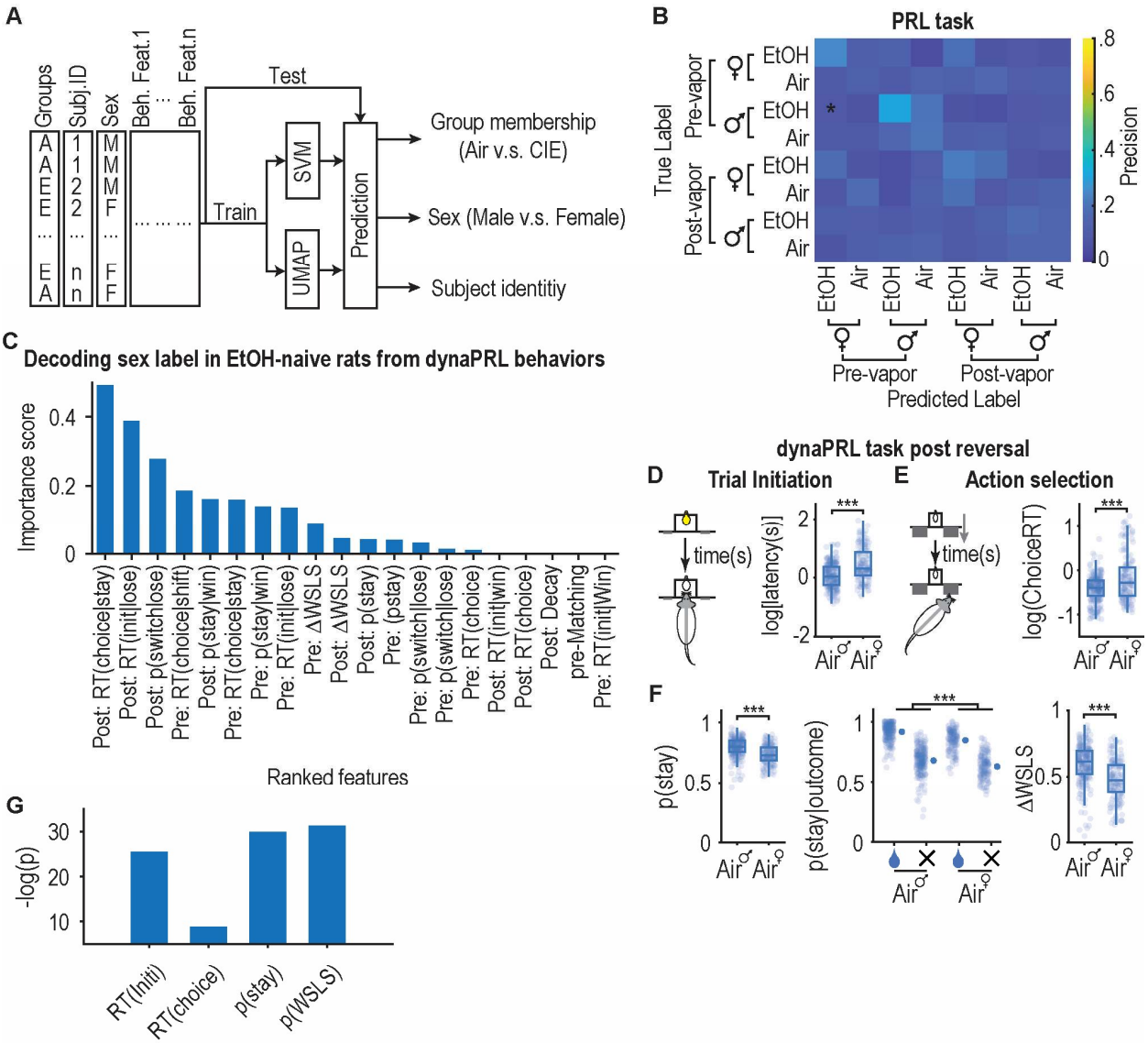

Figure S2. EtOH-exposed rats do not exhibit a distinct behavioral pattern in the standard PRL task. (**A**) The diagram illustrates the classification analysis process. (**B**) Confusion matrix for the SVM multiclass classifier aggregating from 1000 iterations in decoding group membership (EtOH vs. Air), sex label, and exposure conditions (pre- v.s. post-vapor) for N = 13 EtOH rats (8 male, 5 female) and 14 air controls (9 male, 5 female). Diagonal entries represent accurate predictions, while off-diagonal elements represent misclassifications. Classification precision is computed from [true positive rate/(true positive rate + false positive rate)]. *$p < 0.05$ indicates statistical significance determined through Monte Carlo tests. (**C**) Importance scores for each behavioral measurement (feature) were used in SVM decoding to distinguish sex labels in EtOH-naïve (air only) rats. The features are ranked in descending order based on their importance scores. (**D**) Latency of trial initiation after block switch in male and female airs rats during dynaPRL. ***$p < 0.0001$, Wilcoxon rank-sum test. (**E**) Latency of action selection after block switch male and female air rats during dynaPRL. ***$p = 0.00015$, Wilcoxon rank-sum test. (**F**) Stay probability (left panels), stay probability given reward (water drop icon) or no reward (cross, X, symbol) on the last trial (middle panels), and difference of win-stay lose-switch probability (ΔWSLS; right panels) in male and female air rats during the post-reversal phases of

dynaPRL. Left panel, ***$p < 0.0001$, Wilcoxon rank-sum test. Middle panel, ***$p < 0.0001$, two-way mixed-effect ANOVA. Right panel, ***$p < 0.0001$, Wilcoxon rank-sum test. (**G**) Statistical effect, measured by the negative log p-value, comparing among behavioral measures in D-F.

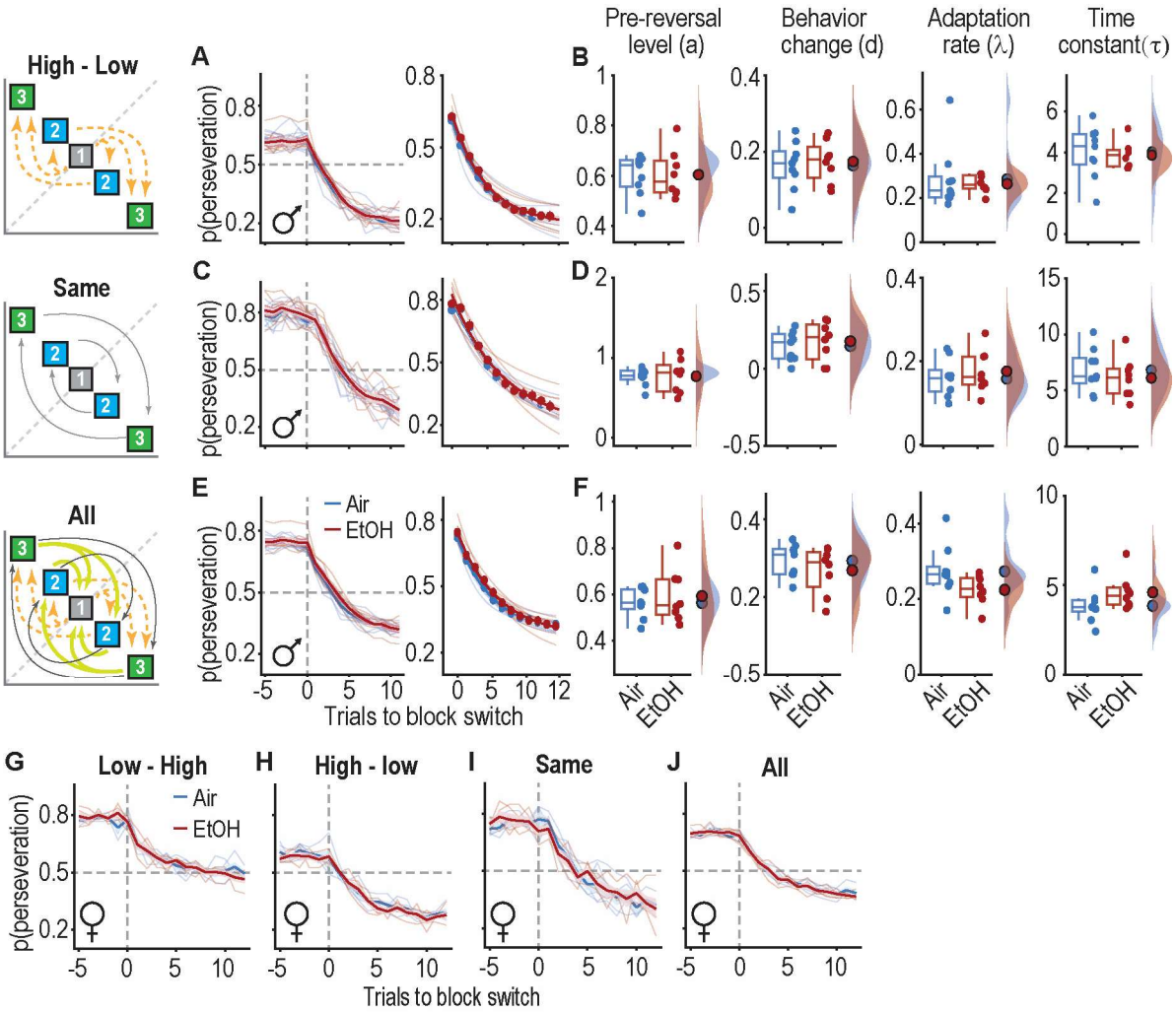

Figure S3. Prior EtOH exposure does not alter reversal learning under low uncertainty in the dynaPRL task. (**A-F**) The probability of selecting the preferred lever before the block switch (left panel) and the performance estimated by the single exponential decay model for learning new action-outcome contingency (right panel) during transitions with High-to-Low uncertainty (A), Same (C), or transitions aggregated across all types (E) in male rats. In the left panels, bold lines indicate the average model fit for the probability of selecting an action based on the pre-switch contingencies. In the right panels, dots represent the true mean data from A, C, and E, and faded lines illustrate model estimates of individual rat choice probability. The estimated parameters' pre-reversal asymptotic value (pre-rev. level), a; magnitude of change in p(preservation) (behavior change), d, adaptation rate, λ, and time constant, τ, for the single exponential decay model during transitions with High-Low (B), Same (D), and transitions aggregated across all types (F) in males. Dots represent individual rats. The box plot depicts the median, 25th, and 75th percentile of the parameter values for each group. Two-way mixed-effect ANOVA for A, C and E. Two-sample independent t test with two-tails for B, D and F. (**G-J**) Probability of selecting pre-switching preferred lever in transitions with Low-High (G), High-Low (H), and Same (I) type transitions as well as transitions aggregated across all types (J) in females. Bold lines indicate the average model fit for the probability of selecting an action based on the pre-switch contingencies. Two-way mixed-effect ANOVA for G-J. N = 13 EtOH rats (8 male, 5 female) and 14 air controls (9 male, 5 female).

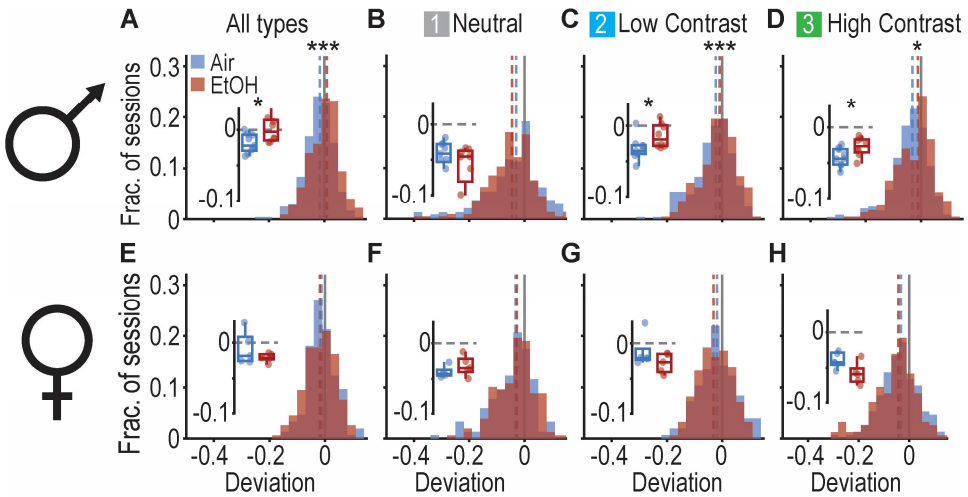

Figure S4. Prior EtOH exposure reduces deviation from matching choice behavior to local reward rate in males but not in females. (**A-D**) Histograms reveal higher matching scores (deviation of choice probability from the reward probability given the choice) in trials from blocks aggregating across all transition types (A), and low contrast (C; 60,30), as well as high contrast (D; 80,10) blocks, but not from neutral blocks (B; 45, 45), in EtOH rats compared to air controls. Vertical blue and red dashed lines represent the median deviation from matching, while the vertical grey solid lines represent perfect matching (0 deviation of choice probability to reward probability given the choice). The inset panels demonstrate the mean deviation score averaged for each EtOH rat and air control, with each dot representing a rat. The horizontal dashed line marks the perfect matching. (**E-H**) Similar to A-D, histograms show no difference in matching scores in all (E), neutral (F), low contrast (G), and high contrast (H) type blocks. *$p < 0.05$ and ***$p < 0.001$. The Mann-Whitney test was used for comparisons in histograms A-H. For all inset panels in A-H, a two-tailed independent two-sample t-test was applied. N = 13 EtOH rats (8 male, 5 female) and 14 air controls (9 male, 5 female).

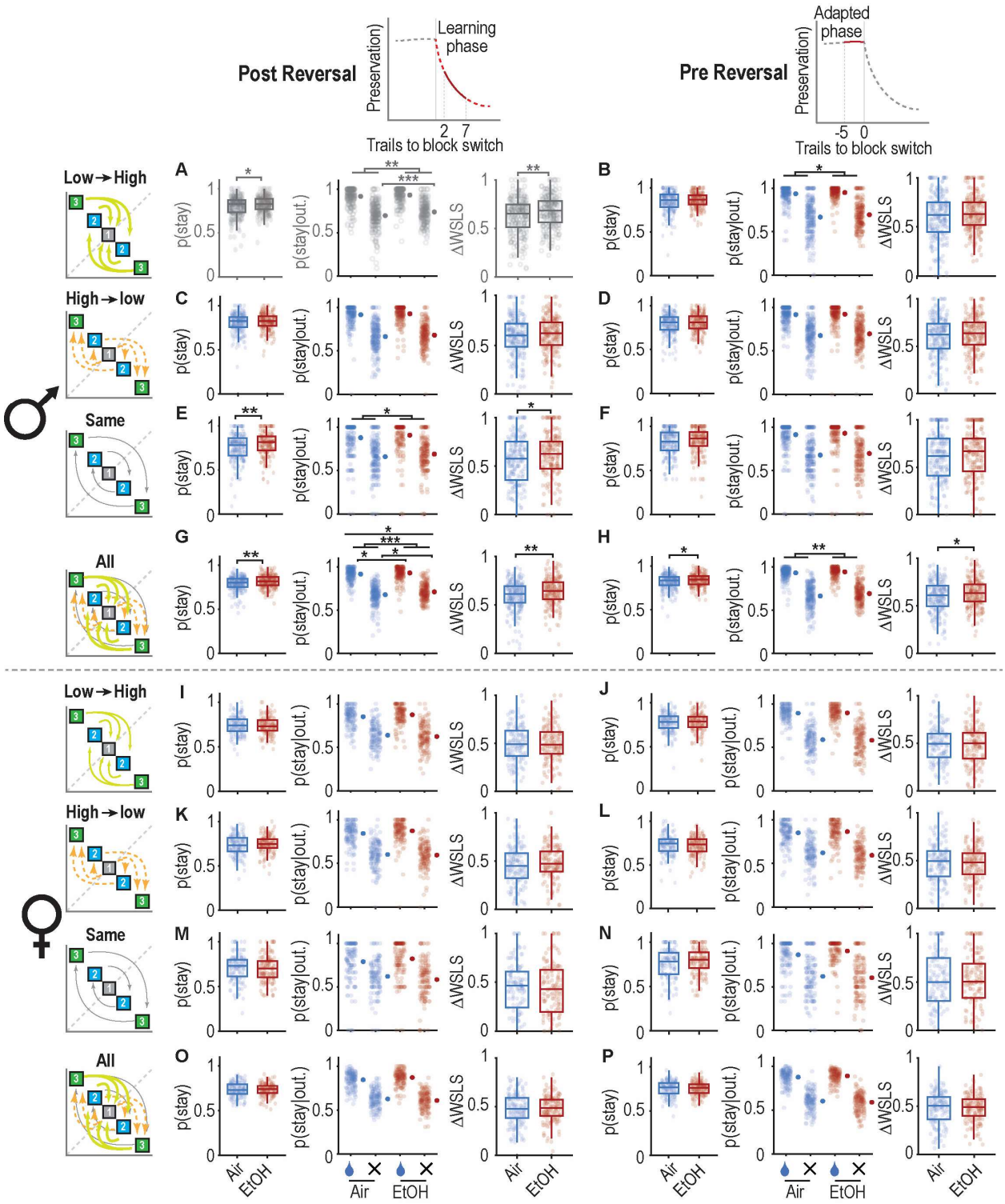

Figure S5. Stay probabilities in pre- and post-reversal under various cognitive challenge conditions. (**A-H**) Stay probability (left panels), Stay probability given reward (water drop icon) or no reward (cross, X, symbol) on the last trial (middle panels), and difference of win-stay lose-switch probability (ΔWSLS; right panels) in the post-reversal (A, C, E, and G) and pre-reversal (B, D, F, and H) phases during various uncertainty conditions in males. Data in A is duplicated from Figures 2H-2J. (**I-P**) Stay probability (left panels), conditional stay probability (middle

panels), and difference of win-stay lose-switch probability (ΔWSLS; right panels) in the pre-reversal (I, K, M, and O) and post-reversal (J, L, N, and P) phases during various uncertainty conditions in females. Dots in all left and right panels represent individual sessions in air and EtOH groups. In all middle panels, dots next to the scatter plot represent the mean of data. Standard errors are vertical lines but are too small to visualize on the graph. *$p < 0.05$, **$p < 0.01$, and **$p < 0.001$, Wilcoxon rank-sum test for the left and right panels in A-P; Two-way mixed-effect ANOVA for the middle panels in A-P. N = 241 sessions from 9 male air rats and 229 sessions from 8 male EtOH rats.

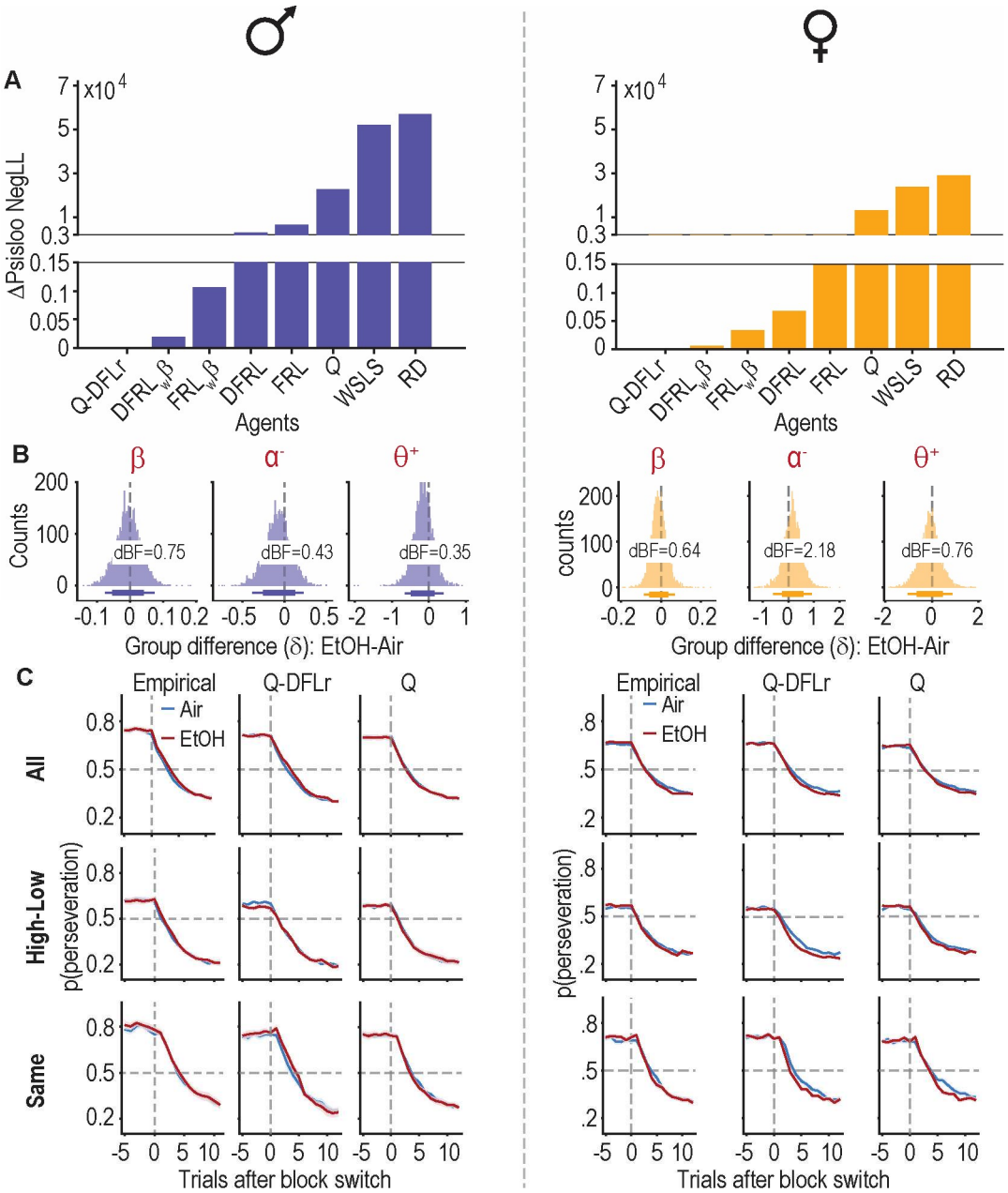

Figure S6. Reinforcement learning models fit trial-by-trial data from the dynaPRL task. (**A**) The bar plot shows the difference of Pareto-smoothed importance sampling (PSIS) with leave-one-out (LOO) negative likelihood (ΔPsisloo NegLL) between candidate models and the best fitting model, Q learning model with differential learning rate (Q-DFLr), on the most left on the x-axis. (**B**) Hamiltonian Markov chain Monte Carlo (HMC)-sampled posterior densities of group difference, δ, representing differences between EtOH- and air-exposed rats in inverse temperature (β in the left panel), negative feedback-mediated learning rates (α^- in the middle panel), and positive feedback-mediated forgetting rates (θ^+ in the right panel), across group-level hyperparameters. Bottom horizontal lines represent the 80% and 95% highest density interval (HDI). dBF: directed Bayes factor. Posterior densities aggregated from 1000 times Monte Carol sampling. (**C**) Choice probability of the initially preferred lever in all, High-low, and Same transitions. Empirical data are reproduced from Figure 1G for male, ♂, and from Figure S3 for

female, ♂, respectively. Simulated data generated using the best-fitting parameters from Q-DFLr and Q models, respectively. Each line represents an average of 20 simulations. Left panels in A-C are data from male (♂) and right panels in A-C are data from female (♀).

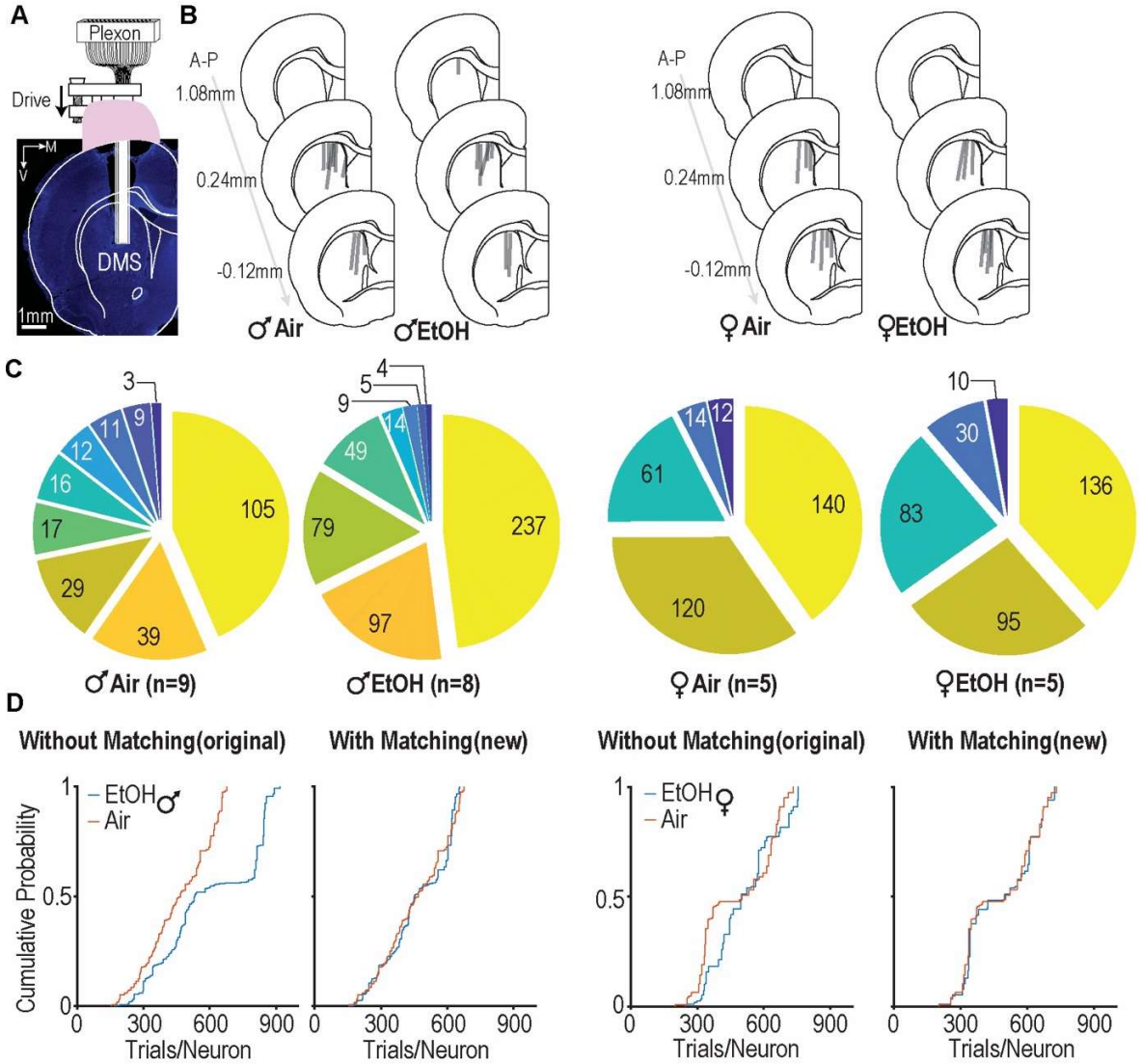

Figure S7. Single unit recording in the dorsomedial striatum (DMS). (**A**) The diagram illustrates a drivable electrode consisting of 16 tungsten wires connected to a Plexon headstage, which is implanted in the Dorsomedial Striatum (DMS). (**B**) A visualization of electrode placement and the range of recording across sessions for each male (♂) and female (♀) animal, denoted by grey lines. (**C**) Pie charts show the aggregate count of neurons recorded from each rat. Chi-squared test. (**D**) he cumulative probability distribution illustrates the number of trials for each recorded neuron before (left) and after (right) trial-matching. This matching was conducted between EtOH and air groups, separately for males and females.

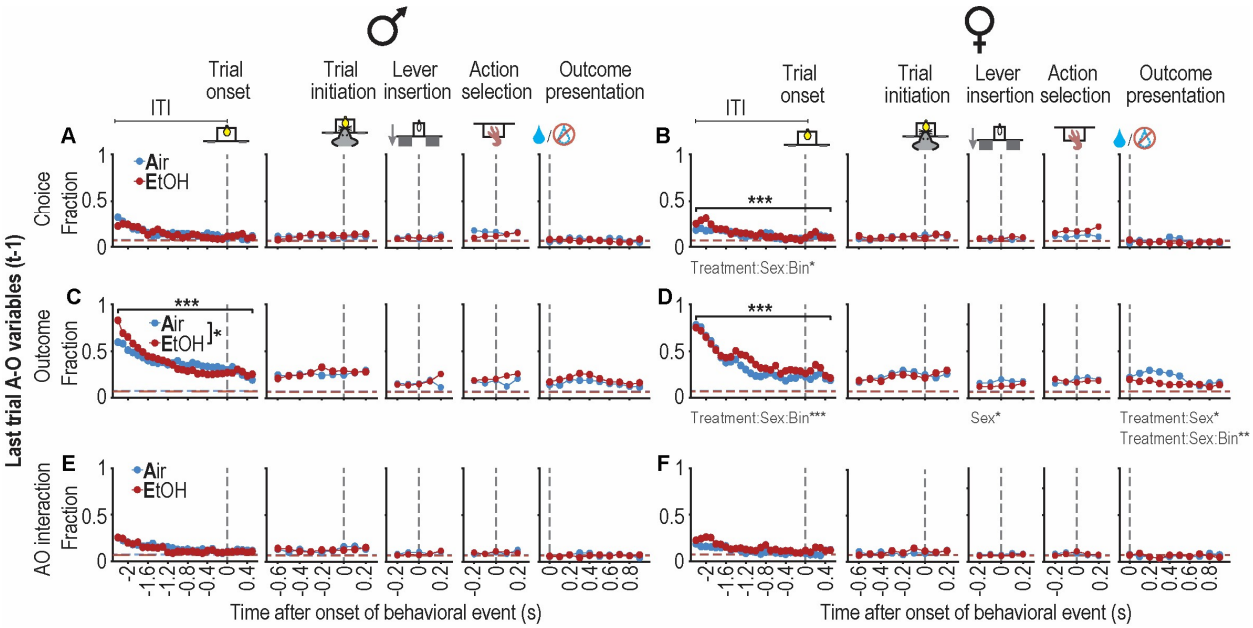

Figure S8. EtOH pre-exposure differentially alters the striatal representation of choice and outcome signals from the previous trial in male and female rats. (**A-F**) Fraction of neurons in males and females showing significant modulation by choice (A for males, B for females), outcome (C for males, D for females), and the interaction of choice and outcome (E for males, F for females) from the previous trial, around behavioral events such as trial onset, trial initiation, lever insertion, lever press, and outcome report. Vertical gray dashed lines mark the onset of each behavioral event. Horizontal blue and red dashed lines indicate the chance level for the air and EtOH groups, respectively. In A, during action selection, $p = 0.018$ main effect of treatment and $p = 0.02$. In B, during trial onset, $p = 0.0007$ interaction between treatment and time and $p = 0.02$ interaction among treatment, sex and time. In C, during trial onset, $p = 0.01$ main effect of treatment and $p < 0.0001$ interaction between treatment and time. In C, during lever insertion, $p = 0.049$ for main effect of treatment and $p = 0.00025$ interaction between treatment and time. In D, during trial onset, $p < 0.0001$ interaction between treatment and time and $p < 0.0001$ interaction between treatment, sex and time. In D, during lever insertion, $p = 0.022$ main effect of sex. In D, during outcome presentation, $p = 0.016$ interaction between treatment and sex and $p = 0.0069$ interaction among treatment, sex and time. The horizontal dashed lines indicate the chance level (binomial test, $\alpha = 0.05$). Logistic regression with Generalized Linear Mixed-Effect Model for A-F. n = 241 neurons from 9 air male rats, 494 neurons from 8 EtOH male rats, 347 neurons from 5 air female rats, and 354 neurons from 5 EtOH female rats.

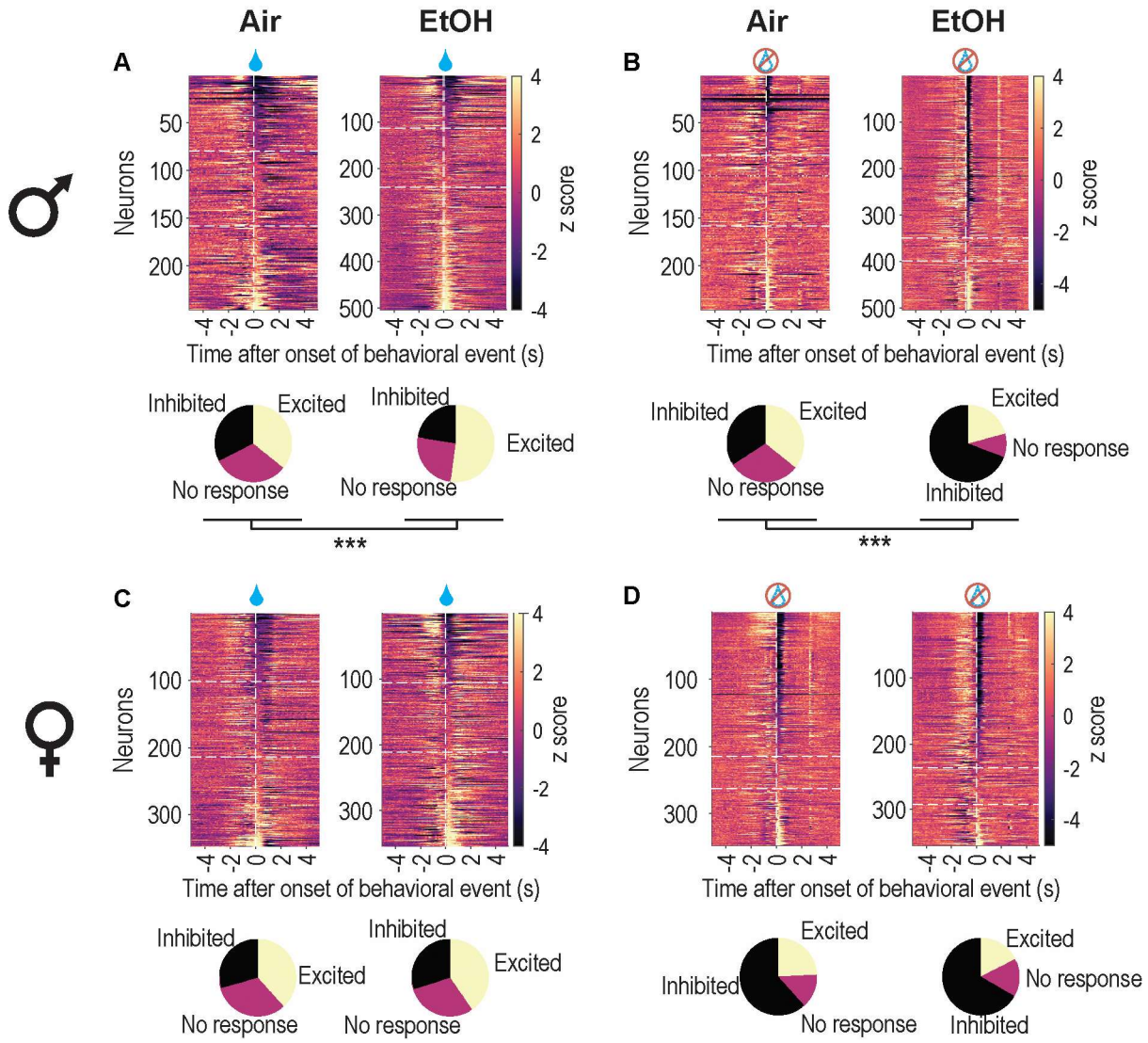

Figure S9. EtOH pre-exposure mainly alters the striatal firing pattern in response to rewarded and unrewarded outcomes. (**A,B**) The top heatmaps illustrate neuronal firing rates from male rats before and after reward delivery (A) or unrewarded outcome presentation (B). ***$p < 0.001$, Chi-squared test. n = 241 units from 9 male air rats, 494 units from 8 male EtOH rats. (**C**,**D**) The top heatmaps illustrate neuronal firing rates from female rats before and after reward delivery (C) or unrewarded outcome presentation (D). Vertical dashed line indicate the onset of reward (A,C) and unrewarded (C,D) outcomes. Neurons are sorted based on their peak z-scored firing rate averaged between 100 ms and 400 ms after the reward (A,C) or unrewarded outcome (B,D) onset. Horizontal lines demarcate subgroups of neurons that are significantly excited (top section), inhibited (bottom section), or non-responsive (middle section) to the reward for A-D. The bottom pie chart depicts the proportion of neurons in each response category for A-D. n = 347 neurons from 5 air female rats, and 354 neurons from 5 EtOH female rats.

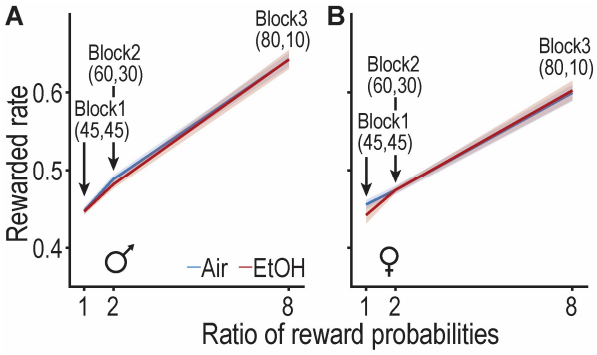

Figure S10. Reward rate as an indicator of block identity. (**A-B**) Reward rates (ratio of rewarded trials to total trials) linearly increase in correlation with the reward probability ratios, ranging from block 1 (45:45) to block 3 (80:10) for both males (A) and females (B). Pearson correlation. N = 13 EtOH rats (8 males, 5 females) and 14 air controls (9 males, 5 females).

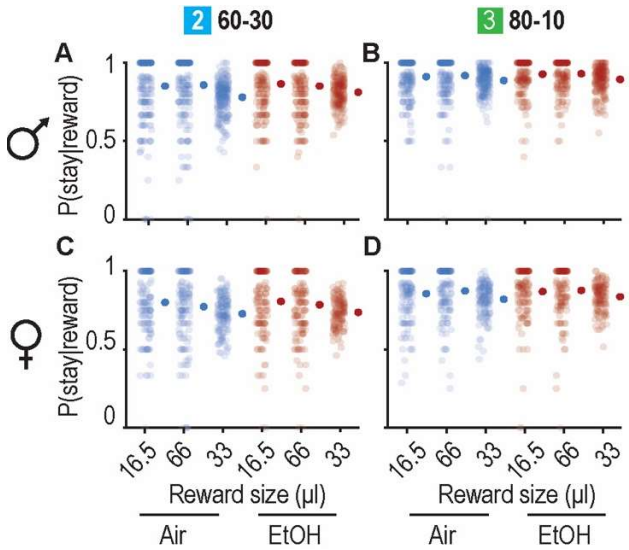

Figure S11. Prior EtOH exposure does not affect rats' response to reward size during the acquired phase in the dynaPRL task. The size of the reward is altered with a 40% chance by doubling or halving the reward size (66 µl or 16.5 µl) compared to the 60% original size (33 µl) 12-trial after switching to block2 (45-45) and block3 (80-10). The probability of win-stay from male rats (A and B) and female rats (C and D), and action selection reaction time (E and F) are significantly modulated by reward size only but not significantly modulated by the interaction of EtOH history and reward size. Main effect of reward, $p < 0.0001$. Dots represent the individual sessions. N = 236 sessions from 9 air male rats, 220 sessions from 8 EtOH female rats, 150 sessions from 5 air female rats and 152 sessions from 5 EtOH female rats. Two-way mixed-effect ANOVA.