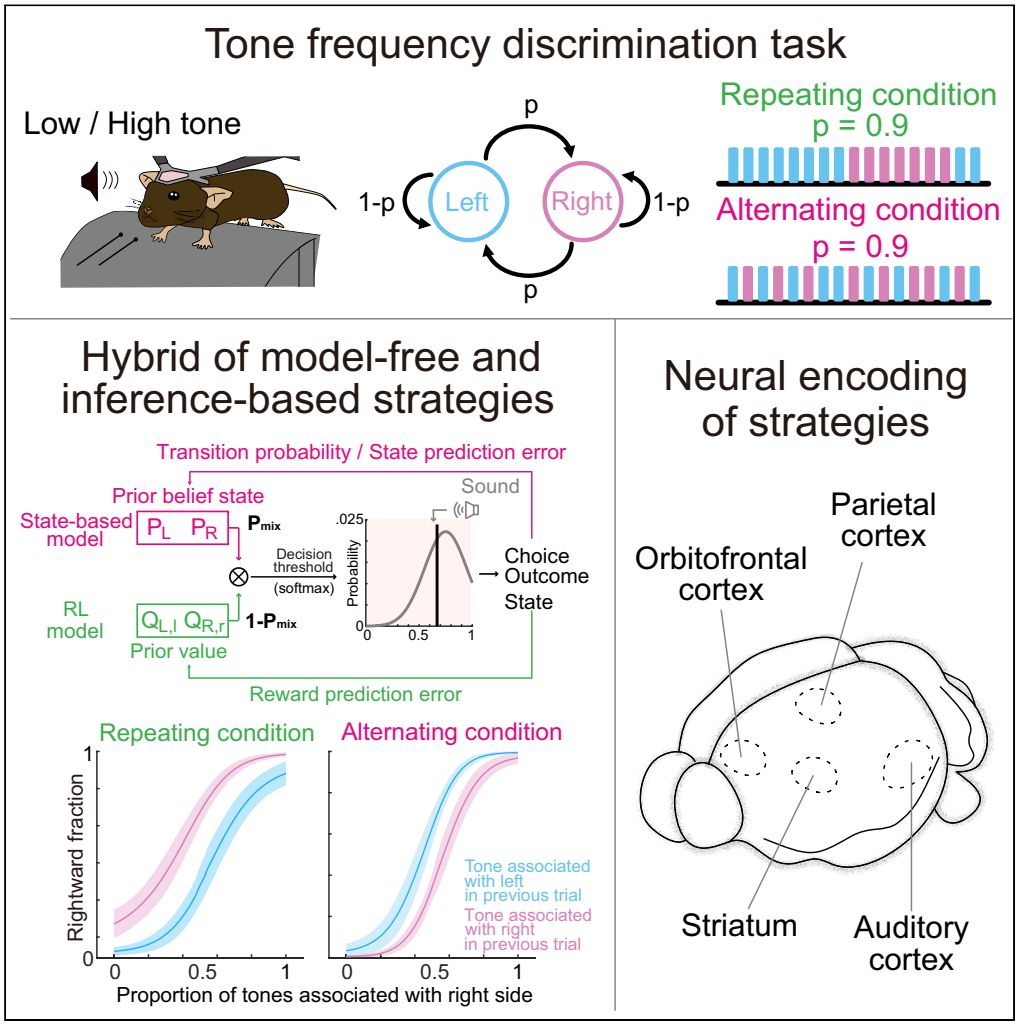
iScience

CellPress
OPEN ACCESS

Article

Global neural encoding of behavioral strategies in mice during perceptual decision-making task with two different sensory patterns



Shuo Wang, Huayi Gao, Yutaro Ueoka, Kotaro Ishizu, Akihiro Funamizu

funamizu@iqb.u-tokyo.ac.jp

Highlights

Tone-frequency discrimination task with the probabilistic alternation of tone categories

Mice use value-based model-free reinforcement learning as the default strategy

Increasing reliance on the inference-based strategy during the training

Global neural encoding of reward expectation in cortical and subcortical regions

Wang et al., iScience 27, 111182
November 15, 2024
https://doi.org/10.1016/j.isci.2024.111182

iScience

CellPress
OPEN ACCESS

Article

Global neural encoding of behavioral strategies in mice during perceptual decision-making task with two different sensory patterns

Shuo Wang,[1,2] Huayi Gao,[1,2] Yutaro Ueoka,[1] Kotaro Ishizu,[1] and Akihiro Funamizu[1,2,3,*]

SUMMARY

When a simple model-free strategy does not provide sufficient outcomes, an inference-based strategy estimating a hidden task structure becomes essential for optimizing choices. However, the neural circuitry involved in inference-based strategies is still unclear. We developed a tone frequency discrimination task in head-fixed mice in which the tone category of the current trial depended on the category of the previous trial. When the tone category was repeated, the mice continued using the default model-free strategy, as well as when the tone was randomly presented, to bias choices. In contrast, when the tone was alternated, the default strategy gradually shifted to a hybrid of model-free and inference-based strategies, although we did not observe distinct strategy changes. Brain-wide electrophysiological recording suggested that the neural activity of the frontal and sensory cortices, hippocampus, and striatum was correlated with the reward expectation in different task conditions, suggesting the global encoding of multiple strategies in the brain.

INTRODUCTION

Perceptual decision-making requires estimating a hidden context from the observation of sensory inputs. Signal detection theory (SDT) shows that to optimize behavior, subjects also need to infer expected outcomes in each context (value) and how the context changes over time (context transition probability).[1,2]

One simple strategy for optimizing behavior by estimating the expected outcome (value) is to estimate and update the value of each choice through trial and error from past direct experiences. This is accomplished by model-free reinforcement learning (RL).[3–5] Model-free RL does not estimate the transition of context when making choices.[3] However, as noted in previous theoretical[6] and experimental research,[7–9] model-free RL is not always the best strategy for optimizing choices. For example, when contexts have certain dependencies or structures, a simple RL model involving only value estimation fails to optimize choices.[10–12] In such complex environments, a behavioral strategy in which the hidden structure of context relationships is inferred becomes important.[13] This is also supported by the SDT, as both value and context estimations are essential for optimizing behavior.[2] A strategy using an internal model of context-transition probability is named the abstract state-based model[7] or inference-based strategy.[8] Recent study shows that, in the early phase of training, mice use model-free RL as a default strategy. Mice then change to an inference-based strategy at the late phase of training when the simple default strategy does not provide sufficient outcomes in a foraging task.[8] Other studies show that mice can switch several behavioral strategies even within a session to optimize choices.[14,15] In this study, we try to investigate how various brain regions, including the frontal and sensory cortices and subcortical regions, represent the different behavioral strategies.

Previous experiments in rodents, monkeys, and humans have shown that the cortico-basal ganglia circuit, which includes the striatum (STR), motor cortex, prefrontal cortex, and sensory cortex, is involved in model-free RL.[4,16–22] In contrast with the model-free strategy, the neural circuit of the inference-based strategy is still under investigation. Early human studies utilizing sophisticated behavioral tasks have identified parallel pathways for model-free and inference- or model-based strategies in the brain,[23,24] while others have shown overlapping involvement of brain regions in these two strategies.[25] These studies revealed that the prefrontal cortex is involved in inference-based strategies.[23–25] Recently, rodent studies have shown that the orbitofrontal cortex (OFC) and hippocampus (HPC) are necessary for the inference strategy,[8,26] while there are a series of studies showing the distributed encoding of task variables across brain areas.[27–29] There is also a report that different behavioral strategies are multiplexed in the motor cortex.[15] Although the neural circuits involved in the inference-based strategy are gradually being identified in some brain regions in animal experiments, it is unclear whether various regions in the brain represent the different behavioral strategies in a distributed manner[15,25,27] or distinct parallel pathways.[6,24]

[1]Institute for Quantitative Biosciences, the University of Tokyo, Laboratory of Neural Computation, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan
[2]Department of Life Sciences, Graduate School of Arts and Sciences, the University of Tokyo, 3-8-2, Komaba, Meguro-ku, Tokyo 153-8902, Japan
[3]Lead contact
*Correspondence: funamizu@iqb.u-tokyo.ac.jp
https://doi.org/10.1016/j.isci.2024.111182

Here, we updated our previous tone frequency discrimination task[1,30–32] to test the neural representations of different behavioral strategies. The task probabilistically alternated the tone category of the current trial based on the category in the previous trial with a transition probability of p.[33,34] We first trained all the mice in the neutral condition (p = 0.5), where there was no bias of tone presentation in the task. We found that although the optimal behavior was an unbiased selection of the left or right choice, mouse behavior was already biased by the outcome in previous trials, suggesting that the default strategy of mice was value-based model-free RL. We then divided the mice into two groups: one group repeated one tone category (repeating condition: p = 0.2), while the other group alternated the tone category in every trial (alternating condition: p = 0.9). Interestingly, the acquisition of proper choice biases was faster in the repeating condition than in the alternating condition. Biased behavior in the repeating condition was achieved from the first session, suggesting that the default model-free strategy was used to optimize choices. In contrast, the acquisition of choice biases in the alternating condition took 3 sessions to achieve, and the behavior was gradually fit to a hybrid model combining model-free and state-based models, with increasing reliance on the state-based model, although we did not find clear separation of behavioral strategies between the task conditions. Value updating was observed even at the overtrained phase in the repeating condition, while the choice was stable in the alternating condition.

We obtained brain-wide electrophysiological recordings from the OFC, HPC, STR, primary motor cortex (M1), posterior parietal cortex (PPC), and auditory cortex (AC) during the overtrained phase of the task. We found that, in both conditions, the neurons in all the recorded regions showed increased activity when the choice was expected to have a high reward probability.[29] In contrast, at the outcome timing, the neurons increased the activity with unexpected outcomes mainly in the repeating condition, possibly because the behavior in the alternating condition was already stable during electrophysiology and did not need to update the choices based on outcomes. These results suggest the global encoding of multiple strategies in the brain.

RESULTS

Mouse choices depend on the transition of tone category in a tone frequency discrimination task

In our tone frequency discrimination task, mice were head-fixed and placed on a treadmill[1,28,30] (Figure 1A, top). Each trial began with retracting the spouts away from the mouse. After a random interval of 1.0–2.0 s, a tone stimulus with a duration of 0.6 s was presented from a speaker placed to the right front of the mouse (Figure 1A, bottom left). The tone stimuli were tone clouds, which were mixtures of low-frequency (5–10 kHz) and high-frequency (20–40 kHz) pure tones (Figure 1A, bottom right).[1,30–32] Depending on the dominant frequency, tone clouds were categorized as low or high. In addition, the tone clouds were named easy (0% and 100% of high-frequency tones), moderate (20%/80% or 25%/75% of high-frequency tones), or difficult (35%/65% or 45%/55% of high-frequency tones). The association between the tone category (low or high) and the correct choice was determined for each mouse. A correct or incorrect choice resulted in the provision of 10% sucrose water (2.4 μL) or a noise burst (0.2 s), respectively.

In our task, the tone category of the current trial was probabilistically alternated based on the tone category in the previous trial with a transition probability of p (Figure 1B).[33,34] We first exposed all the mice to the neutral condition (p = 0.5), in which the tone categories were randomly presented. We then divided the mice into two groups. In the repeating condition (p = 0.2), tone categories frequently repeated across trials, while in the alternating condition (p = 0.9), tone categories alternated across trials. Both in the repeating and alternating conditions, the first 40 trials only contained 100% low- or high-tone clouds with fixed tone sequences (STAR Methods). We did not use the first 40 trials for analyses.

In the neutral condition, although the tone category was randomly selected in each trial, the choice behavior of the mice was biased toward the side that was rewarded in the previous trial (Figures 1C and S1A), suggesting that the mice had a default strategy to repeat the previously rewarded choice. We then analyzed the choices of mice in the repeating and alternating conditions (repeating condition, 113 sessions in 8 mice; alternating condition, 141 sessions in 11 mice). In the example repeating condition, the mouse chose the right side more frequently after trials in which right-side rewarded tones were used than after trials in which the left-side rewarded tones were used, indicating repeating choice biases (Figure 1F). On the other hand, in the alternating condition, the mouse tended to switch choices. To quantify choice bias, we analyzed how the choice in the current trial depended on the tone category in the previous trial. When the task condition switched from the neutral to the repeating condition, the mice immediately showed repeating choice bias from the first session. In contrast, mice in the alternating condition required 3 sessions on average to acquire the alternating biased behavior (Figures 1G, S1D, and S1E). Since mice had repeating choice biases in the neutral condition, these results suggest that mice continuously used the same default strategy to optimize choices in the repeating condition, while mice required some sessions to switch the choice biases in the alternating condition. In the neutral, repeating, and alternating conditions, a logistic regression analyzed how the correct and error choices in the past trials affected the choice in the current trial[34] (STAR Methods). We found that the previous correct trials had the largest influence on the choice in the current trials in all the conditions (Figures 1D, 1E, 1H–1J, and S1C). We also found that the previous correct trials had a larger influence than the previous error trials in the repeating condition (Wilcoxon signed-rank test, p = 0.0078 and 0.32 in the repeating and alternating conditions) (Figure S1F).

Mice have different strategies between repeating and alternating conditions

We investigated the behavioral strategies of mice under the repeating and alternating conditions. As the first 40 trials only contained the easy tone clouds, we did not include the first 40 trials in the analyses (STAR Methods). We first proposed two models to analyze mouse choices based on the SDT, which predicted that estimations of expected outcomes and hidden-context probabilities are essential for optimizing behavior.[1,2] Model-free RL estimated the expected choice outcome in the low- and high-tone-category states, while the belief probabilities of right- and left-rewarded states were constant. The state-based model estimated the probability of the current context as the prior belief

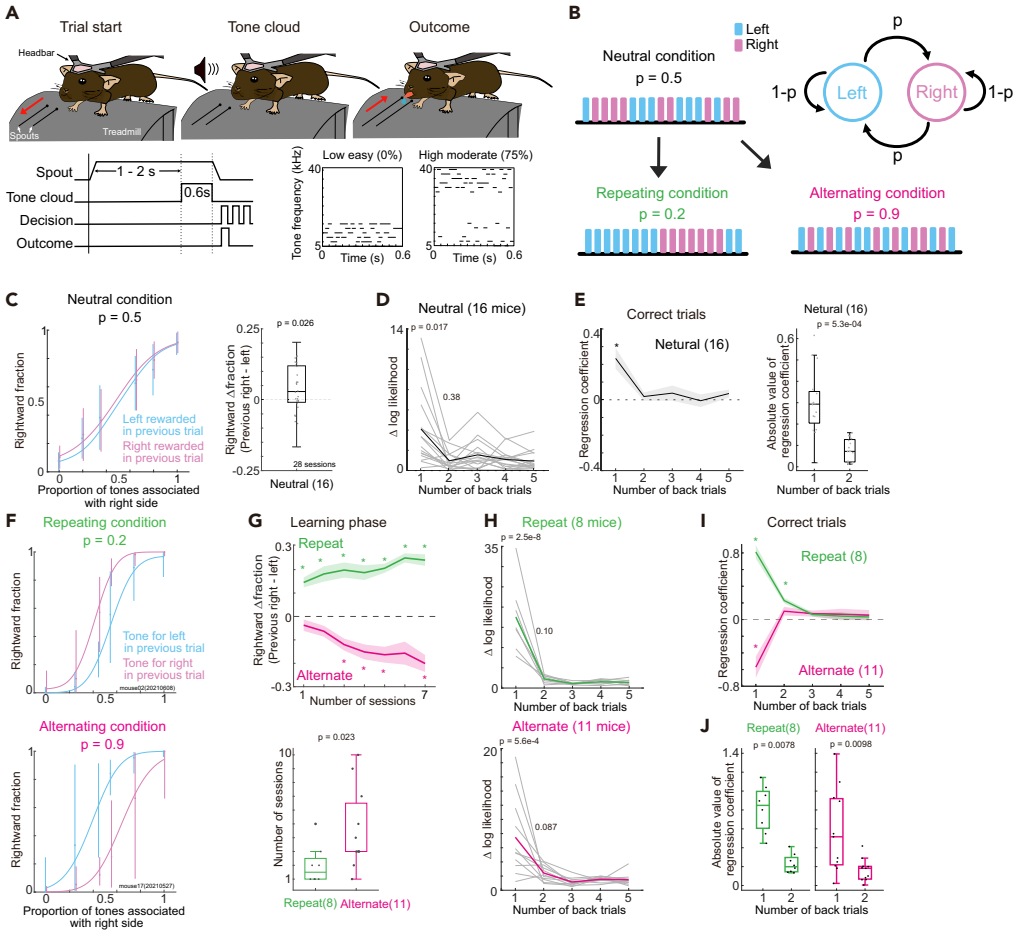


Figure 1. Tone frequency discrimination task in head-fixed mice with repeating and alternating conditions

(A) Task scheme. Each trial started by moving the spouts away from the mouse. After a random interval of 1–2 s, a sound stimulus (tone cloud) was presented from the speaker positioned in the right front of the mouse. The spouts immediately approached the mice after the end of the sound. Mice licked either the left or right spout to receive sucrose water. The right panels show example tone clouds. During the overtrained electrophysiological phase, the spouts were moved 0.5 s after the end of the sound.

(B) Task conditions. The task included neutral, repeating, and alternating conditions. After the neutral condition with a transition probability of $p = 0.5$, mice were exposed to either the repeating condition ($p = 0.2$) or the alternating condition ($p = 0.9$).

(C) Psychometric function in the neutral condition ($p = 0.5$) (28 sessions in 16 mice). Data are represented as mean ± SD. (right) Mice significantly biased their choices to the previously rewarded side (linear mixed-effects model) (central mark in the box: median, edges of the box: first quartile (Q1) and third quartile (Q3); bars: most extreme data points without outliers, here and hereafter).

(D) Logistic regression analyzed how the past events affected the choices in the current trial in the neutral condition. The likelihood ratio test examined whether the additional events from 1- to 5-back trials improved the prediction accuracy of the current choice (Δ log likelihood). Average model fitting is shown. Black and gray lines show the mean Δ log likelihood in all the mice and in each mouse, respectively.

(E) Regression coefficients in the analysis of (D) for the correct choices. *$p < 0.01$ in the Wilcoxon signed rank test. Data are represented as mean ± SEM (left). Comparisons of absolute regression coefficients for the 1-back and 2-back correct trials in the neutral condition (right). p value in the Wilcoxon signed rank test. 28 sessions of 16 mice in the neutral condition. Boxplots are the same as C.

(F) Example psychometric function of choice behavior in the repeating (left) and alternating conditions (right). Data are represented as means and 95% confidence intervals.

(G) Comparison of the number of sessions required to achieve the proper choice biases in the repeating and alternating conditions after switching from the neutral condition. (top) Rightward Δ fraction was the difference in the average fraction of right-side choice after the right- and left-rewarded trials. Data are represented as mean ± SEM (repeating condition: 56 sessions in 8 mice; alternating condition: 77 sessions in 11 mice, *$p < 0.01$ in the Wilcoxon signed rank test). (bottom) The number of sessions required to achieve the proper choice biases (STAR Methods) (8 and 11 mice; Mann–Whitney U test).

(H) Past events affected the current choices in the repeating and alternating conditions. Average model fitting is shown. The colored and gray lines show the mean Δ log likelihood in all the mice and in each mouse, respectively.

(I) Regression coefficients in the analysis of (H) for the correct choices. *$p < 0.01$ in the Wilcoxon signed rank test. Data are represented as mean ± SEM. 8 and 11 sessions in the repeating and alternating conditions.

(J) Comparisons of absolute regression coefficients for the 1-back and 2-back correct trials in (H). p value in the Wilcoxon signed rank test. Boxplots are same as C.

state (Figure 2A). The prior belief state was estimated from the previous state and the state-transition probability, which was updated by a state prediction error in each trial (STAR Methods).[35] We also proposed a hybrid model which had a weighted combination of the model-free RL and state-based model (Figure 2B).[14,15] In example sessions, the simulated choices generated by the model-free RL, and state-based, hybrid, f-memory models captured the repeating and alternating choice biases in the repeating and alternating conditions, respectively (Figures 2C, 2D, and S1K).

We first analyzed the choices in the neutral condition. We found that the RL model matched the mouse choices better than the state-based model did, suggesting that the mice used a model-free strategy as the default strategy (Figures 2E and S1B). To robustly assess the strategy of mice, we also introduced a memory-based model-free RL model (memory strategy) reported in a previous study[34] with small changes for our study (STAR Methods). We found that the RL model and the hybrid model fit the mice choices better than the memory strategy in the repeating and alternating conditions, respectively (Figure 2F). We then updated the memory strategy[34] by introducing a forgetting value updating,[17,28] named the forgetting memory strategy (f-memory strategy) (STAR Methods).

In the repeating condition, the RL model fit the choice behavior better than all the other models, suggesting that mice continuously used a simple model-free strategy, as well as in the neutral condition (Figure 2F). The RL model was a more consistent fit for choice behavior than the hybrid model beginning from the first session (Figure 2I). In contrast, in the alternating condition, the hybrid and f-memory models equally matched the mice choices better than the RL model did (Figure 2F, bottom and 2G). The hybrid model gradually fit the choices, and the weight of the state-based model increased by experiencing more sessions (Figures 2H and 2I). Also, the state-based model gradually fits the mice choices compared to the RL model (Figure S1G). In addition, simulated choices by the f-memory model captured the effects of past events in the neutral and repeating conditions, while the hybrid model captured the effects in both the repeating and alternating conditions but not in the neutral condition (Figures 2J, 2K, S1L, and S1M). These results suggest that while the default model-free strategy was used to optimize choices in the repeating condition, in the alternating condition, mice gradually increased their reliance on the inference-based strategy.

We analyzed behavior in the overtrained phase, in which we electrophysiologically recorded the neural activity of the mice. We analyzed 178 sessions in 14 mice during the overtrained phase. In the repeating condition (51 sessions in 5 mice), when the number of correct repeated choices increased, the choice biases increased (Figures 3A and 3B, top). This finding suggested that the reward expectation was updated even during the overtrained phase. In contrast, in the alternating condition (127 sessions in 9 mice), the alternating choice biases did not depend on the number of alternated correct choices (Figures 3A and 3B, bottom). Similar to the learning phase, logistic regression showed that the previous correct trials had the largest influence on the choice in the current trial in both conditions (Figures 3C, 3D, S1H, and S1I). The events in the 2-back trial slightly differently affected the current choice between the two task conditions (Figure S1H, likelihood ratio test, p = 0.046 and 0.12 in the repeating and alternating conditions). These differences potentially contributed to the difference in choice biases across conditions (Figure 3A).

In the repeating condition, the RL model fits the mice choices than all the other models (Figure 3E, top). In contrast, in the alternating condition, the state-based, hybrid, and f-memory models fit the choices more than the RL model (linear-mixed effects model, p = 8.1e-7–7.1e-4) (Figure 3E, bottom and S1J). The hybrid and f-memory models similarly fit the mice's choices in the alternating condition. The direct comparison of model fitting (Figure 3F), the model fitting in individual sessions (Figure 3G), and the effects of past events in simulated choices in the hybrid model (Figures 3H and S1N) suggested that mice relatively used the inference strategy in the alternating condition compared in the repeating condition.

Choice and reward are widely encoded in the brain, and only the auditory cortex encodes sound

To compare the neural encoding of different strategies in the repeating and alternating conditions, we used a Neuropixels 1.0 probe to electrophysiologically record neural activity at the overtrained phase of the task (Figure 4A). The transition probability of the tone category was set to p = 0.2 and p = 0.8 in the repeating and alternating conditions, respectively. We targeted the OFC, PPC, HPC, and AC in both the repeating and alternating conditions and additionally recorded the activity of M1 and STR in the alternating condition (Figures 4A and S2). We used one Neuropixels probe in each session and analyzed the activity of 14 mice in 178 sessions. After spike sorting, we identified 27668 neurons in total (STAR Methods).

We first detected 12749 task-relevant neurons that exhibited significantly increased activity compared to the baseline activity in one of the 70 time windows during the task (p < 1.0e-10 in the one-sided Wilcoxon signed rank test; 46.08% of all recorded neurons; repeating condition, 3599 out of 9514 neurons (37.83%); alternating condition, 9150 out of 18154 neurons (50.40%)) (Figures 4B, S3A, and S3B). The duration of each window was 0.1 s between −1.5 and 2.5 s from sound onset (40 windows). The time windows were also set between −0.5 and 2.5 s from the choice timing (30 windows; 40 + 30 = 70 windows in total). The baseline for sound-onset activity was defined as the activity at −0.2–0 s from the start of the trial, i.e., spout removal. The baseline for choice activity was defined as the activity at −0.2–0 s from the time the spout approached, i.e., between the end of the sound and the choice. Among the task-relevant neurons, we targeted the neurons that showed increased activity (1) before the sound (−0.6–0 s from the sound onset), (2) during the sound (0–0.6 s from the sound onset), and (3) during the choice and outcome (0–1.0 s from the choice) (STAR Methods). The maximum false discovery rate of all the task-relevant neurons in all the brain regions was 4.1e-10 (MATLAB, mafdr) (Figure S3C). To confirm the temporal sequence of task-relevant neurons, we randomly split the trials in each session into half and averaged the activity of each neuron in each half of the trials. We then analyzed the maximum activity timing of neurons in each half of the trials and analyzed the Spearman correlation between the timings.[28] We repeated this procedure 100 times to reduce noise

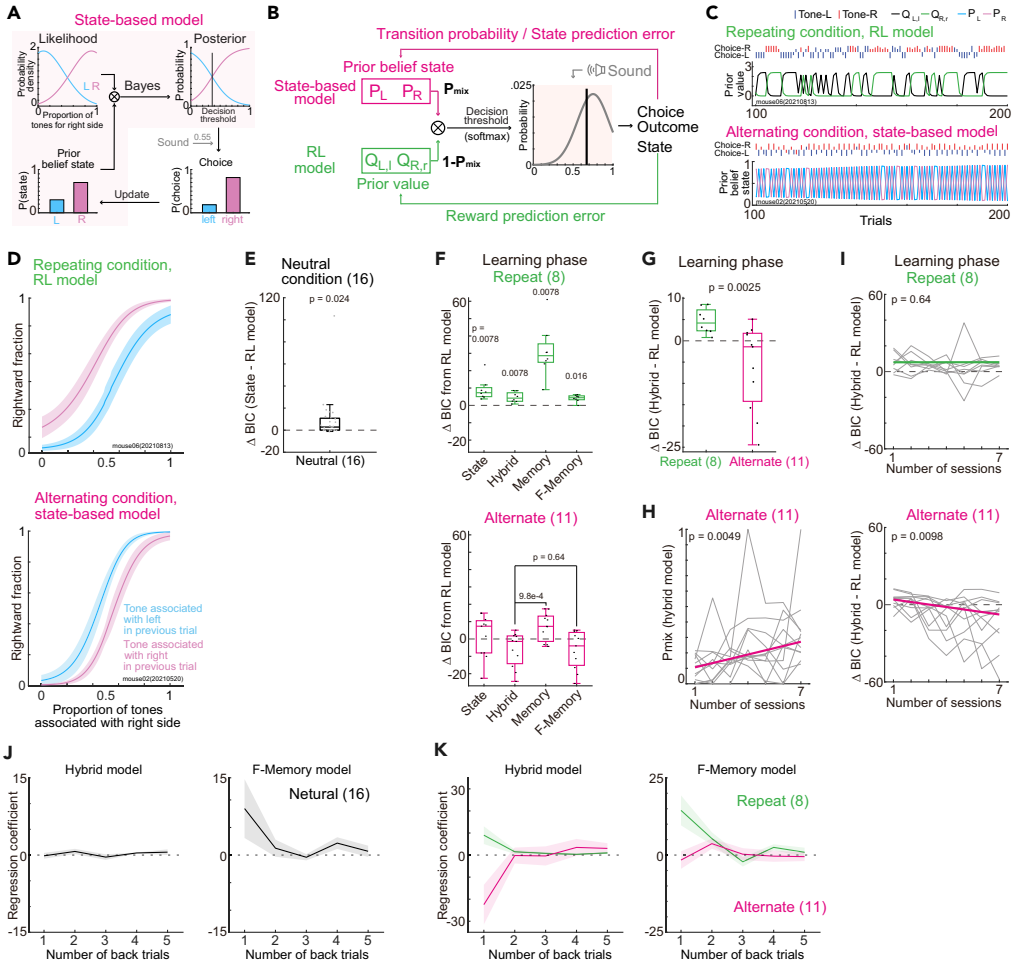


Figure 2. Mice used different strategies in the repeating and alternating conditions

(A) Scheme of the state-based model. Bayesian inference computed the decision threshold of choice based on the prior belief state and the sensory distributions.

(B) Scheme of the hybrid model. The hybrid model integrated the model-free and state-based models with a weight parameter P_{mix} to decide choices.

(C) Left and right prior values ($Q_{L,l}$ and $Q_{R,r}$) were estimated from the RL model in an example session of the repeating condition (top). Prior belief states (P_L and P_R) were estimated from the state-based model in the alternating condition (bottom).

(D) Example sessions with simulated choices with the RL model in the repeating condition (top) and with the state-based model in the alternating condition (bottom). We simulated the mice choices 100 times based on the fitted parameters in the RL and state-based models. Data are represented as mean ± SD.

(E) Model fitting in the neutral condition. Δ BIC (Bayesian information criterion) was the difference in fitting between the state-based and RL models (28 sessions in 16 mice, linear mixed-effects model). Boxplots are the same as Figure 1C.

(F) Model fitting of the state-based, hybrid, memory, and f-memory models compared to that in the RL model during the learning phase (8 and 11 mice, *p* value in the Wilcoxon signed rank test). Boxplots are the same as E.

(G) Comparison of Δ BIC between the repeating and alternating conditions during the learning phase. Δ BIC was the difference in model fitting between the hybrid and RL models (8 and 11 mice, *p* value in the Mann−Whitney U test). Boxplots are same as E.

(H) Ratio of using the state-based model in the hybrid model (P_{mix}) in sessions 1 to 7. We tested whether the slope of P_{mix} was significantly negative or positive (*p* value in the Wilcoxon signed rank test, 11 mice). Gray line shows the mean Pmix in each mouse. Magenta line shows the slope of regression analysis.

(I) Difference in model fitting between the hybrid and RL models (Δ BIC) in sessions 1 to 7. We tested whether the slope of Δ BIC was significantly negative or positive (*p* value in the Wilcoxon signed rank test, 8 and 11 mice in the repeating and alternating conditions). Gray line shows the mean Δ BIC in each mouse. Colored line shows the slope of the regression analysis.

(J) Regression analysis with simulated choices with the hybrid and f-memory model in the neutral condition. We simulated the mice choices 100 times based on the fitted parameters. Regression coefficients of 1- to 5-back trials are shown. Data are represented as mean ± SEM. 28 sessions in 16 mice.

(K) Regression coefficients of 1- to 5-back trials simulated by the hybrid and f-memory model in the repeating and alternating condition. Data are represented as mean ± SEM. (8 and 11 mice).

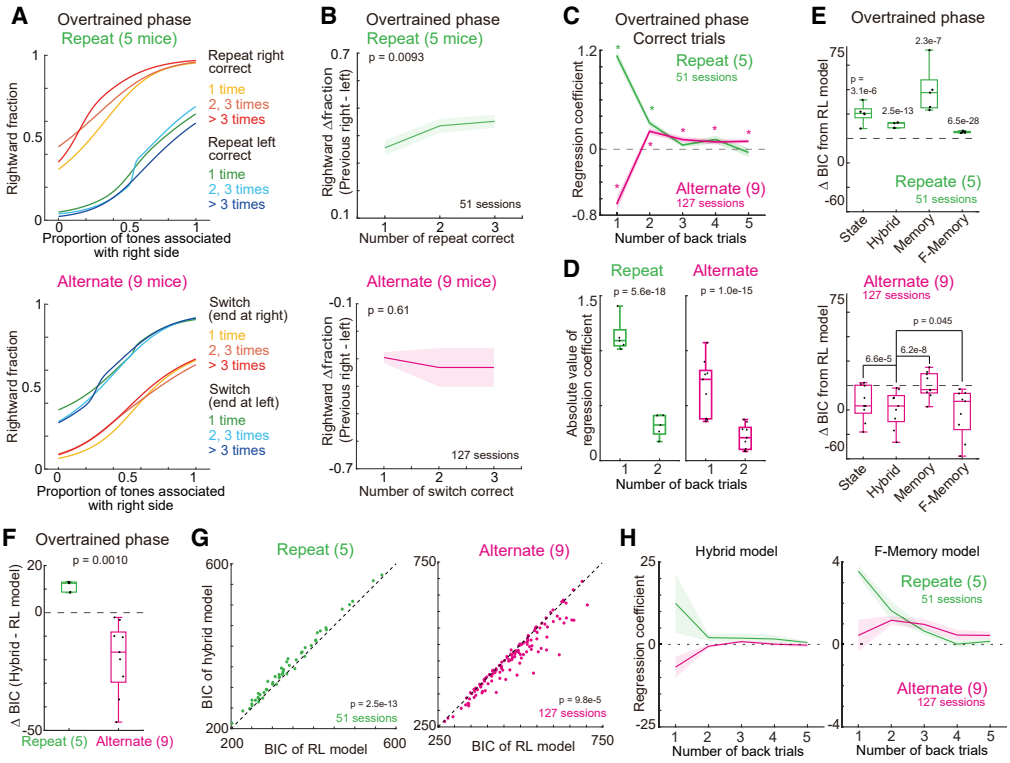


Figure 3. Choice behavior at the overtrained phase during electrophysiology

(A) Mean psychometric functions of choice behavior in the repeating (51 sessions in 5 mice) and alternating conditions (127 sessions in 9 mice) at the overtrained phase. The transition probability of the tone category in the alternating condition was set to $p = 0.8$.

(B) Rightward Δ fractions increased when the number of repeated correct choices increased in the repeating condition. Rightward Δ fractions remained stable in the alternating condition (178 sessions in 14 mice in total, p value in the linear mixed effects model). Data are represented as mean ± SEM.

(C) Regression coefficients of the logistic regression analysis investigating the effects of previous correct trials on the current choice. Data are represented as mean ± SEM. *$p < 0.01$ in the linear-mixed effects model (51 and 127 sessions from 5 to 9 mice in the repeating and alternating conditions, respectively).

(D) Comparison of the absolute regression coefficients in (C) for the 1-back and 2-back correct trials in the repeating (left) and alternating conditions (right). p value in the linear mixed effects model. Boxplots are the same as Figure 1C.

(E) Model fitting of the state-based, hybrid, memory, and f-memory models compared to that in the RL model during the overtrained phase (p value in the linear mixed effects model). Boxplots are the same as D.

(F) Comparison of Δ BIC (Bayesian information criterion) between the repeating and alternating conditions (5 and 9 mice, p value in the Mann−Whitney U test). Boxplots are the same as D.

(G) Comparison of model fitting between the RL and hybrid models in each session (51 and 127 sessions from 5 to 9 mice in repeating and alternating conditions, p value in the linear-mixed effects model).

(H) Regression analysis with simulated choices with the hybrid and f-memory model in the repeating and alternating condition. Regression coefficients of 1- to 5-back trials are shown. Data are represented as mean ± SEM (51 and 127 sessions from 5 to 9 mice in the repeating and alternating conditions).

from the random grouping of trials. We found that the average Spearman correlations of maximum activity timings ranged between r = 0.56 and 0.82 in all the recorded regions (Figures 4B, S3A, and S3B).

Here, we analyzed the neurons with increasing activity during the task as the uniform criterion across regions, similar to our previous studies.[28,36] In the prefrontal cortex, a previous study showed that the increasing neurons were modulated by task variables than the decreasing neurons.[37] Sensory modulations were analyzed only with the neurons with increasing neurons.[38,39] In this study, we confirmed that the proportion of decreasing neurons in all the recorded brain regions was smaller than that of increasing neurons (Figure S3D: decreasing and increasing neurons in repeating condition: 6.36–11.30% and 26.12–40.06%; alternating condition: 6.08–15.71% and 33.54–62.34%). We also confirmed using the regression analysis that the number of decreasing neurons representing task variables was smaller than those of increasing neurons (Figures S3E and S3F).

We used a generalized linear model (GLM) with 10-fold cross-validation to investigate the neural encoding of task variables, including choices, sounds, and outcomes, in the current or previous trials in addition to running speed (Figure 4C) (STAR Methods). As the previous choices had the largest influence on the current choices in the overtrained phase (Figure 3D), we included the previous events in the analysis. We first validated whether our GLM captured the neural encoding of task variables by comparing the deviance of GLM fitting between the recorded and shuffled neural activity (Figure 4D). We found that our GLM in the recorded neural activity had

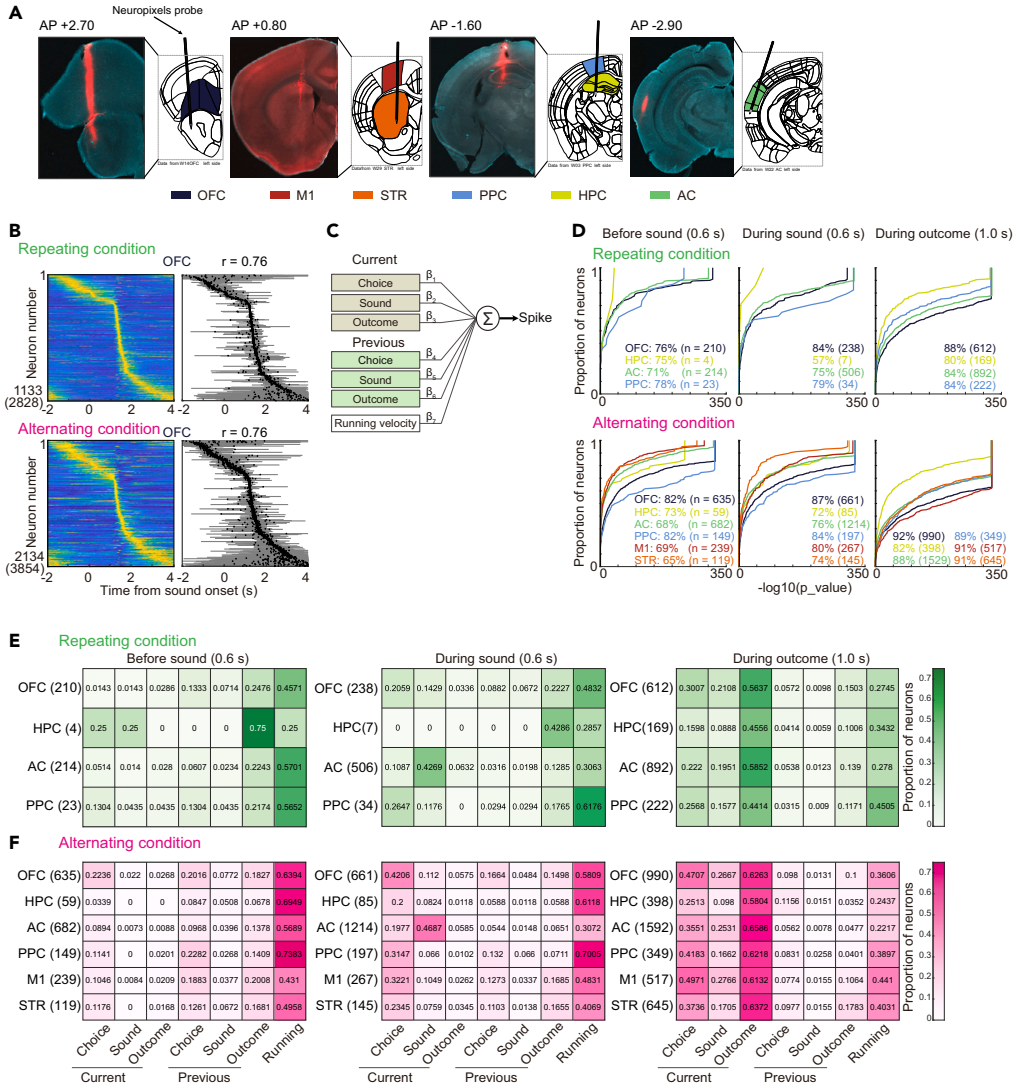


Figure 4. Choice, sound, and outcome representations in the OFC, M1, STR, PPC, HPC, and AC

(A) Neuropixels 1.0 probe traces for the OFC, M1, STR, PPC, HPC, and AC in example mice. The red color shows the location of the Neuropixels 1.0 probe.

(B) Cross-validated average activity of task-relevant neurons in the OFC in the repeating (top) and alternating conditions (bottom) ($p < 1.0e-10$ in the one-sided Wilcoxon signed rank test). (left) Activity of each task-relevant neuron in half of the trials was normalized between 0 and 1, sorted by the maximum activity timing based on the other half of the trials. The parentheses show the number of all the recorded neurons. (right) Validation of temporal sequence of task-relevant neurons with cross validation. Black dot and gray bar show the mean ± SD of the maximum activity timings in each task-relevant neuron (Spearman correlation, r = 0.76 and 0.76).

(C) Schematic of the generalized linear model (GLM) for testing neural encoding.

(D) Validation of GLM. The deviance of GLM with recorded neural activity was compared with the distribution of deviance in GLM in the shuffled activity. The deviance of GLM in shuffled activity was analyzed 200 times. The inset shows the percentage of task-relevant neurons that had a significantly lower deviance than that in the distribution of shuffled GLM ($p < 0.025$).

(E and F) Proportion of neurons representing the sounds, choices, and outcomes at previous and current trials. We analyzed the proportion of neurons with GLM with 10-fold cross validation. Numbers in the parentheses show the number of the subset of task-relevant neurons that showed a significant increase in activity during each time window.

smaller deviance than that in the shuffled activity in many task-relevant neurons ($p < 0.025$) (repeating condition: 57–88%, alternating condition: 65–91%).

We then investigated the proportion of neurons representing each task variable (Figures 4E and 4F). In this analysis, when the deviance of full-parameter GLM was lower than the mean – 1.96 times the standard deviation of deviance in the one-parameter-removed GLM, we defined that the neuron represented the removed task variable (STAR Methods).[40] Before sound onset, the proportions of neurons in

each brain region representing the choice in the previous trial were 6.07–13.33% and 8.47–22.82% in the repeating and alternating conditions, respectively. The proportions of neurons representing previous outcomes were 21.7–75.0% and 6.78–20.1%. The previous choice and outcome were similarly represented in some brain regions (chi-square test: repeating condition: OFC, AC, p = 0.0072–7.4e-6; HPC, STR, p = 0.083–0.48; alternating condition in all the brain regions: p = 0.027–0.76), consistent with the behavioral findings that the previous correct choices affected the choice in current trial (Figures 3C and 3D). During the sound, the proportion of sound-representing neurons in the AC was greater than that in the other regions (AC: 42.69 and 42.87% in the repeating and alternating conditions; OFC, PPC, HPC, M1, STR: 0–14.29%; p = 1.7e-44–7.6e-8 in the chi-square test). At the time of choice, all 6 brain regions were more likely to represent the outcome than the choice or sound (chi-square test; outcome vs. choice: all 6 regions: p = 1.9e-61–0.012; outcome vs. sound: all 6 regions: p = 1.7e-99–5.0e-16). These results suggest that the neurons in the AC represent choices, sounds, and outcomes, whereas the neurons in the other brain regions represent choices and outcomes.

Previous studies report that facial movements are represented in neurons as well as the task variables.[41–43] Among the 178 sessions, we recorded the facial movements in 88 sessions from 6 mice (repeating and alternating conditions: 11 and 77 sessions in 1 and 5 mice). We captured the mice's faces with one camera in front of the mice with a sampling rate of 140 Hz (STAR Methods) (Figures S4A and S4B). From the movie data, we extracted the 9 facial features with DeepLabCut (DLC)[44] and computed the motion strength by summing the absolute velocities of 9 features. We found that, before and during the sounds, facial motion strengths were mainly correlated with previous outcomes, while they were correlated to the events in a current trial at the outcome timing (Figure S4C).

We therefore compared the proportion of neurons representing task variables between the GLM with and without facial motion strengths (STAR Methods) (Figure S4D). In the GLM with facial movements, the proportion of neurons representing task variables decreased (Wilcoxon signed rank test in the 6 brain regions and 7 task variables (6 x 7 = 42 in total in each time window): before sound: p = 2.0e-4; during sound: p = 0.016; during outcome: p = 1.1e-4), although the overall distribution of neural encoding was similar (Figure S4D). These data suggest that the facial movements are represented in the task-relevant neurons.

Neurons in wide brain regions modulate activity according to previous choices, but the auditory cortex represents previous sounds and choices in the alternating condition before the sound onset

We first analyzed neural activity before sound onset (Figure 5A). We identified 2259 neurons that exhibited a significant increase in activity between −0.6 and 0 s from sound onset (451 and 1883 neurons in the repeating and alternating conditions, respectively; 12.53% and 20.58% of the task-relevant neurons). In the repeating condition, the example neuron in the OFC showed increased activity when the previous trial was a left-rewarded choice followed by a left choice in the current trial (Figure 5B, top). In the alternating condition, an OFC neuron showed increasing activity in the previous right-rewarded trials and the current left-choice trials (Figure 5B, bottom). Thus, example neurons in the OFC showed increased activity in response to the correct choice in previous trials before sound onset (Figure 5B).

To confirm the neural representations of previous choices, we compared the choice indices between the previous and current trials (STAR Methods). The choice index compared activity during low- and high-category sound trials and ranged between −1 and 1.[28,45] By comparing the absolute choice indices for the previous correct and current trials, we quantified whether the neurons encoded the previous or current choice (Figure 5C). We found that the previous choice indices in correct trials in all the brain regions were greater than the choice indices in the current trial, suggesting that the neurons in multiple brain regions represented the previous choice (Figures 5C and S5A).

To simultaneously analyze the activity of previous left- and right-choice representing neurons, we defined the preferred side of each neuron based on the choice index (STAR Methods) (Figure 5A). For the neural activity before sound, the choice index was analyzed based on the choice in the previous trial. In the repeating condition, left-side preferred neurons had a larger activity in the previous left- than previous-right-correct trials, whereas right-side preferred neurons had a large activity in previous right-correct trials. In contrast, in the alternating condition, as mice were required to switch choices in 80% of trials, we reversed the definition of neurons from the repeating condition: the left- and right-side preferred neurons had a large activity in the previous-right- and previous-left-correct trials, respectively. These neurons had significantly different activities between the previous left- and right-correct choice trials ($p < 0.01$ in the Mann−Whitney U test).

In the alternating condition, we could not clearly separate whether mice used the f-memory model or hybrid model for choices (Figure 3). The f-memory model required memorizing the previous choice to compute the values, while the inference-based strategy in the hybrid model required memorizing the previous sound category as the hallmark of the true state (STAR Methods). We, therefore, investigated whether the neural activity represented the previous choice or sound by comparing the choice indices before sound presentation, following previous correct and incorrect trials. When neurons represented sound categories, the correlations of the choice indices between correct and incorrect trials were negative. Conversely, when neurons represented choices, the correlations were positive.[28,45] In the repeating condition, the choice indices of neurons in the OFC and AC were positively correlated between after previous correct and incorrect trials, suggesting the previous choice representations (Figures 5D and 5E, top). In contrast, in the alternating condition, although the neurons in OFC represented the previous choice, the neurons in AC showed a negative correlation in a linear regression analysis (β = −0.027) (Figures 5D and 5E, bottom). To confirm whether the AC changed the representation between the repeating and alternating conditions, we investigated whether the choice indices between the two conditions were modeled with either one identical linear regression or two different regressions. Although the choice index in the OFC fit to the one regression (Bayesian information criterion (BIC) in one and two regressions: −1465 and −1459), the choice index in the AC fit to the two independent regressions (BIC: −828 and −839).

The hybrid model parameterized the ratio of model-free and inference-based strategies (STAR Methods, Figure 2B). We found that the ratio of the inference-based model varied across sessions (Figure 5F). We thus divided the sessions into half by the median ratio (0.36) and

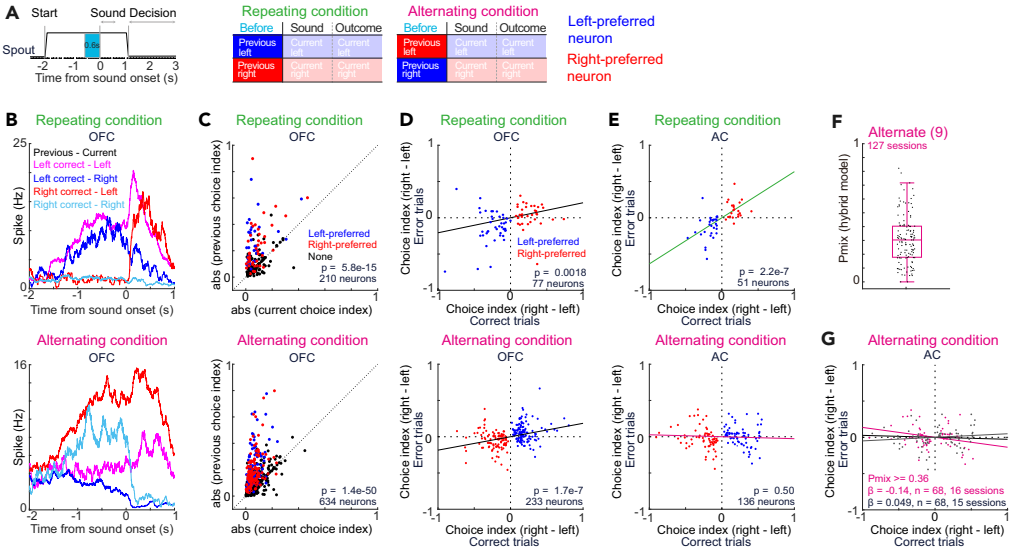


Figure 5. Neural activity before sound onset depends on previous events

(A) Analysis of neural activity between −0.6 and 0 s from sound onset. Right panels show the definition of left- and right-preferred neurons before sound.

(B) Average activity of example neurons from OFC before sound onset.

(C) Scatterplot comparing the absolute value of choice indices for previous and current trials in the OFC (STAR Methods). The previous choice indices in correct trials were higher than the current choice indices, suggesting that the neurons represented previous choices (blue and red dots: left- and right-side preferred task-relevant neurons, $p < 0.01$ in the Mann−Whitney U test; black dots: non-side-preferred task-relevant neurons; p value in the Wilcoxon signed rank test).

(D) Choice indices of task-relevant neurons in the OFC. The choice indices of OFC neurons had positive correlations between the correct and incorrect trials in the previous trials, suggesting the previous choice representation. p values show the significance of regression coefficients. One linear regression was fit to the choice indices in both the repeating and alternating conditions (Bayesian information criterion (BIC) in one and two regressions: −1465 and −1459). The activity of left- and right-side preferred task-relevant neurons is shown. Black lines show the slope of regression analysis.

(E) Choice indices of task-relevant neurons in the AC. Two independent linear regressions were fit to the choice indices in the repeating and alternating conditions (BIC in one and two regressions: −828 and −839). Colored lines show the slope of regression analysis.

(F) Ratio of the inference-based strategy in the hybrid model at the overtrained phase in each session of the alternating condition. Boxplot is the same as Figure 1C.

(G) Comparison of the choice-index slope between the sessions with high and low ratios of the inference-based strategy. The sessions were categorized based on the median ratio (0.36). The number of sessions was different from (F), as we only analyzed the sessions with AC recordings. The inset shows the regression coefficients, number of neurons, and number of sessions. Magenta and black lines show the slopes of the regression analysis for sessions with high and low ratios of inference-based strategy, respectively.

independently analyzed the regression coefficients in the choice index (Figure 5G). The regression coefficient was significantly negative in the sessions with the high ratio of inference-based strategy ($\beta = -0.14$, $p = 0.031$). These results suggest that the AC is a candidate brain region for representing inference-based strategy.

Neurons in wide brain regions modulate activity according to upcoming choices before sound onset

Next, we investigated whether the neural encoding of previous choices was modulated by the upcoming choice in the current trial (Figure 6A). In the repeating condition, when the previous choice was the preferred choice, the OFC neurons significantly increased the activity in current trials when mice repeated the choice rather than when they switched their choice even before the sound was presented (Figure 6A, top left; Figure 6B). Conversely, in the alternating condition, the OFC neurons increased the activity when the mice switched from the previous choice (Figure 6A, bottom left). When the previous choice was the nonpreferred side, neurons in both the repeating and alternating conditions showed opposite activity compared to the preferred previous choice (Figure 6A, right). The neurons in the AC, PPC, STR, and M1 showed activity patterns similar to those of the OFC (Figures 6C, 6D, S5B, and S5C). These global neural modulations, which depending on the choice sequence, were similarly observed for the current upcoming correct or incorrect choices (Figures S5D and S5E), and the modulations were larger in the incorrect trials. In the PPC in repeating condition, we observed only 4 neurons which had significantly different activities between the left- and right-correct choice in previous trials ($p < 0.01$ according to the two-sided Mann−Whitney U test). We thus did not show the result in the PPC (Figure 6C).

As the facial movements were represented in the task-relevant neurons (Figure S4D), we used a GLM to quantify whether the neural modulations depending on current choices (Figures 6C and 6D) were independent of the facial motion strengths (STAR Methods). The GLM investigated whether the neural activity before sound was correlated to the choices and facial motion strengths. We found that the facial motion strengths did not affect the overall results of neural modulations in the alternating condition (Figure 6E). These results suggest that the neurons in multiple brain regions not only represented the previous choice but also modulated the activity according to the upcoming choice.

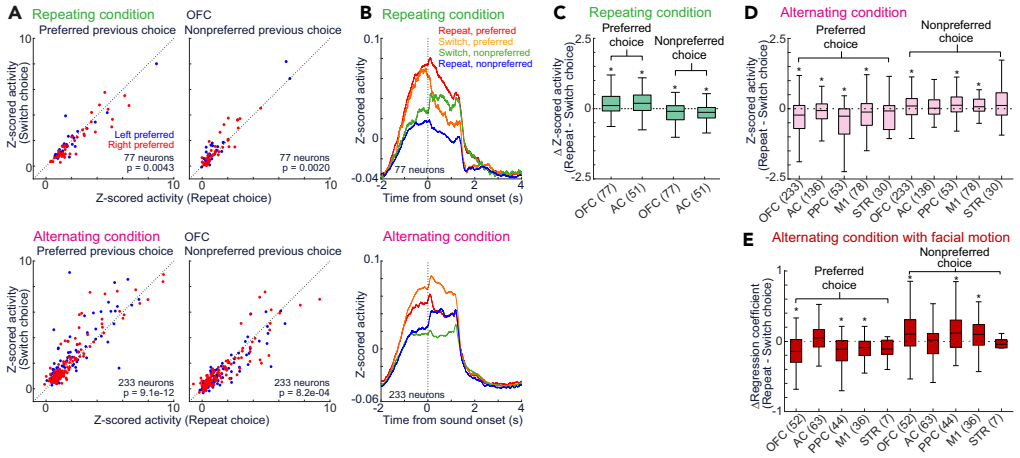


Figure 6. Neural activity before sound onset depends on upcoming choices

(A) Average activity during repeated choices (x axis) and switched choices (y axis) between −0.6 and 0 s from sound onset (*p* value in the Wilcoxon signed rank test). The activity of left- and right-side preferred task-relevant neurons is shown. The definition of left- and right-preferred neurons was same as Figure 5A.

(B) Average traces of neurons in the OFC in a current trial before sound onset. Neurons were identical to (A). The traces were categorized based on the definition of preferred neurons in Figure 5A. Data are represented as mean.

(C and D) The difference in average activity between the repeated and switched choices before sound onset. Because we only detected 4 side-preferred task-relevant neurons from the PPC in the repeating condition, we did not show the data from the PPC. Boxplots are the same as Figure 1C, but outliers, defined as values beyond 1.5 times the interquartile range, were excluded from the plots. The numbers in parentheses show the number of left- and right-side-preferred neurons. *$p < 0.05$ in the Wilcoxon signed rank test.

(E) Correlation between the neural activity and choice sequences in alternating sessions with the facial-movement recording (77 sessions in 5 mice). Regression analysis investigated how the neural activity correlated to the preferred and non-preferred choices and the facial motion strengths. Δ regression coefficient in y axis shows the difference in regression coefficients between the repeated and switched choices. The parentheses show the number of left- and right-side-preferred neurons. *$p < 0.05$ in the Wilcoxon signed rank test. Boxplots are the same as C and D.

The activity of neurons during sound presentation depends on the choice sequence

During the sound presentation, we identified 785 and 2535 neurons in the repeating and alternating conditions, respectively, that exhibited a significant increase in activity (21.81% and 27.79%, respectively, of the task-relevant neurons) (Figure 7A). In the repeating condition, the activity of the example neuron in the OFC gradually increased when the previous trial was right-rewarded, followed by a right choice in the current trial (Figure 7B, top). In the alternating condition, an example OFC neuron increased the activity when the choice was switched from left to right (Figure 7B, bottom). For the neural activity during sound, we defined the preferences of neurons based on the choice in the current trial (Figure 7A). The choice indices of neurons in the OFC, PPC, STR, HPC, and M1 were positively correlated between the correct and incorrect trials, suggesting the choice encoding (Figures 7C and S6A). In contrast, the choice indices of AC neurons were negatively correlated with each other, suggesting that they were involved in sound encoding (Figure S6A).

We investigated whether the activity of neurons was changed by the choice sequence (Figure 7D). In the repeating condition, the OFC neurons showed increased activity when the preferred choice was repeated compared with when the choice was switched to the preferred side (Figure 7D, top left; Figure 7E). In contrast, in the alternating condition, the OFC neurons increased the activity when the choice was switched to the preferred side (Figure 7D, bottom left). In the nonpreferred choice, the OFC neurons showed opposite activity than in the preferred choice condition in both the repeating and alternating conditions (Figure 7D, right). Similar to the OFC, the other recorded brain regions exhibited choice-sequence-dependent activity (Figures 7F, 7G, S6B, and S6C), independent of the facial motions (Figure 7H), similarly between the correct and incorrect choices (Figures S6D and S6E). Given that repeated and switched choices had high reward expectations in the repeating and alternating conditions, respectively, these results suggest that the neurons in various brain regions encode reward expectations in different behavioral strategies.

Global neural encoding of unexpected outcomes in the repeating but not in the alternating condition

At the time of the outcome (Figure 8A), we identified 1895 and 4491 task-relevant neurons in the repeating and alternating conditions, respectively (52.56% and 49.08% of the task-relevant neurons). The preferences of neurons were defined based on the choice in the current trial. An example OFC neuron in the repeating condition showed increased activity when the mouse switched its choice from right to left rather than when it repeated the left choice (Figure 8B, top). In contrast, in the alternating condition, the OFC neuron did not show choice sequence-dependent activity (Figure 8B, bottom). The choice index during the outcome timing suggested that all the recorded regions had choice encoding (Figures 8C and S7A).

We analyzed whether the change in activity of the neurons depended on the previous choice (Figure 8D). In the correct trials in the repeating condition, the OFC neurons exhibited increased activity for the preferred choice when the choice was switched compared with

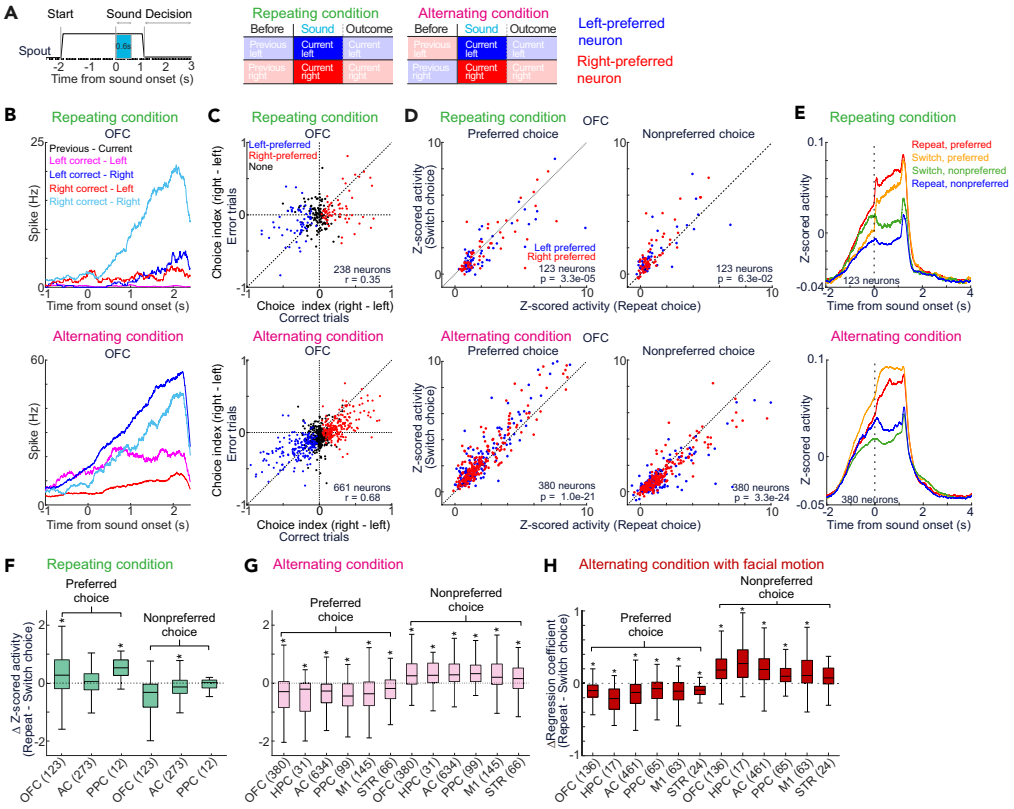


Figure 7. Neural activity during sound presentation

(A) Analysis of neural activity during 0.6 s of sound. Right panels show the definition of left- and right-preferred neurons during sound.

(B) Average activity of example neurons in the OFC during sound.

(C) Choice indices of task-relevant neurons. The choice indices of OFC neurons had positive correlations between the correct and incorrect trials, suggesting choice representation. Data plots are the same as Figure 5.

(D) Average activity of each neuron during the repeated (x axis) and switched choices (y axis) (p value in the Wilcoxon signed rank test).

(E) Average activity of neurons in the OFC. Data are represented as mean.

(F and G) The difference in average neural activity between the repeated and switched choices during sound. The parentheses show the number of left- and right-side-preferred neurons. *$p < 0.05$ in the Wilcoxon signed rank test. Boxplots are the same as Figures 6C and 6D.

(H) Correlation between the neural activity and choice sequences during sounds in the alternating condition with facial-movement recording (77 sessions in 5 mice). Data plots and analyses are the same as Figure 6E. *$p < 0.05$ in the Wilcoxon signed rank test. Boxplots are the same as Figure 6E.

when it was repeated (Figure 8D top left, 8E, 8F). In contrast, in the incorrect trials, OFC neurons exhibited greater activity when choices were repeated than when choices were switched (Figure S8). Given that reward expectations were lower in switched choices than in repeated choices, these results suggest that the neurons in the OFC encode unexpected outcomes. We found similar activity patterns in the AC and PPC (Figures 8F, S7, and S8), suggesting that multiple brain regions globally represented unexpected outcomes for the model-free strategy.

In contrast, in the alternating condition, none of the recorded brain regions except the HPC showed choice sequence-dependent activity in the correct trials (Figures 8G and S7), irrespective of the facial motions (Figure 8H). In the incorrect trials, only the HPC and PPC neurons exhibited a change in activity based on the reward expectation (Figure S8). During the overtrained phase of electrophysiological neural recording, the mice exhibited experience-dependent choice updates in the repeating condition, while their behavior was stable in the alternating condition (Figures 3A and 3B). These results were consistent with the activity at the outcome time, in which the activity was modulated by the reward expectation mainly in the repeating condition.

DISCUSSION

To investigate the neural representation of model-free and inference-based strategies in multiple brain regions, we used a tone frequency discrimination task in head-fixed mice with different transition probabilities in the tone category. We found that mice tended to repeat previously rewarded choices even when the tone category was randomly selected in the neutral condition (Figure 1C). This default strategy was continuously used in the repeating condition to properly bias the choices. In contrast, mice took several sessions to reverse the choice biases in the alternating condition (Figure 1G).

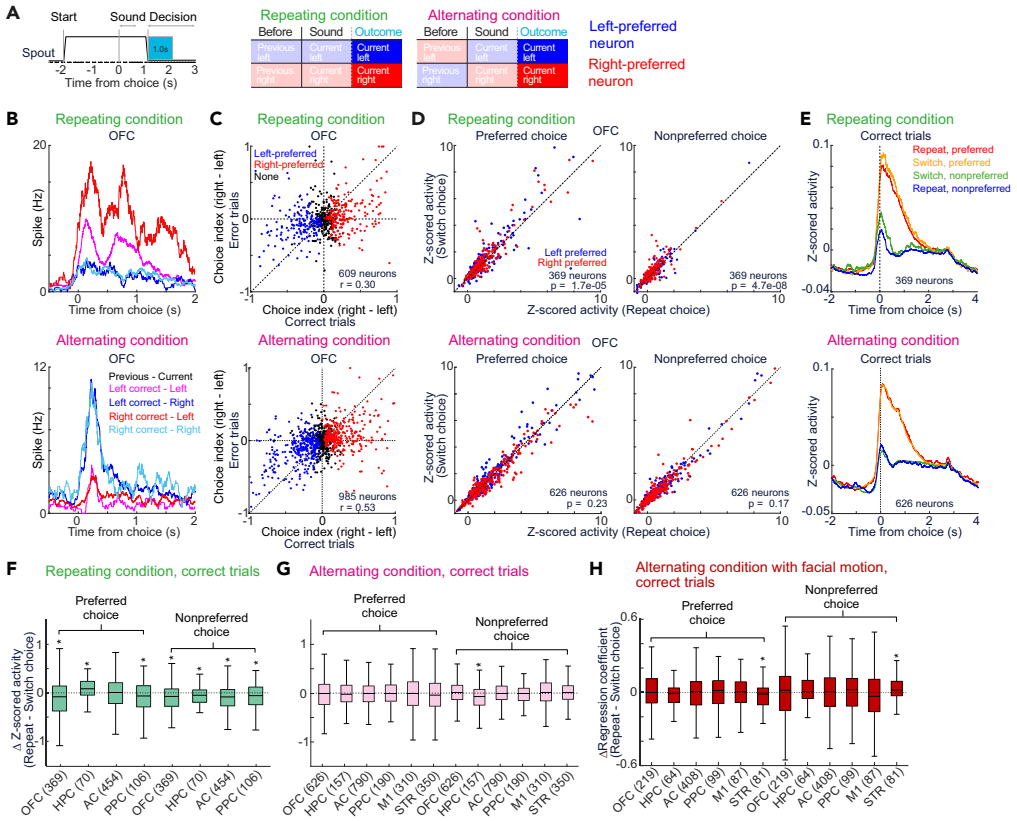


Figure 8. Neural activity at outcome timing

(A) Analysis of neural activity between 0 and 1.0 s from choice and outcome. Data presentations are consistent with those in Figures 5, 6, and 7.

(B) Average activity of example neurons during the outcome including both correct and incorrect trials.

(C) Choice indices of task-relevant neurons. The choice indices of OFC neurons suggested the choice representation (blue, red, and black dots: left-, right-, and non-side-preferred neurons, respectively).

(D) Comparison of activity between the repeated and switched choices (p value in the Wilcoxon signed rank test).

(E) Average activity of neurons in the OFC. Data are represented as mean.

(F and G) The difference in average activity between the repeated and switched choices in the current correct trials. Outliers are not shown in the plots. The parentheses show the number of left- and right-side-preferred neurons. *$p < 0.05$ in the Wilcoxon signed rank test. Boxplots are the same as Figures 6C and 6D.

(H) Correlation between the neural activity and choice sequences in current correct trials in the alternating condition with the facial-movement recording (77 sessions in 5 mice). *$p < 0.05$ in the Wilcoxon signed rank test. Boxplots are the same as Figure 6E.

An RL model estimating the expected outcome of each choice fit the mice's choices in the neutral and repeating conditions, suggesting that the default model-free strategy was used in the repeating condition. In the alternating condition, the mixed strategy of model-free and inference-based, as well as the strategy of model-free memory-based model (f-memory model), fit the mice choices. The mixed ratio gradually shifted from the model-free to inference-based strategy which determined a hidden context by estimating the transition probability of tone category (Figure 2). During the overtrained phase with an electrophysiological recording with Neuropixels, the behavior of the mice was fit to the model-free strategy under the repeating condition (Figure 3E). In contrast, the hybrid strategy and the f-memory model equally fit to the mice choices in the alternating condition. Thus, it was difficult to identify the strategy of mice in the alternating conditions by only observing the behavior.

The inference-based strategy required memorizing the previous sound category, or the true rewarded condition, to decide choices, while the f-memory model memorized the previous choice for value computing (STAR Methods). We therefore investigated whether any brain region represented the previous sound category in the alternating condition, while it represented the previous choice in the repeating condition for model-free RL. We found that, before sound presentations, neurons in the AC, but not in the OFC, shifted the representations from previous choices to previous sounds from the repeating to the alternating condition (Figures 5D and 5E). These results show that the AC is surprisingly a candidate brain region for representing the inference-based strategy.

In general, the neurons in all the recorded regions, including the OFC, PPC, HPC, STR, M1, and AC exhibited increased activity when the preferred choices or tones expected a large reward probability, suggesting that brain-wide encoding of reward expectations occurred both in the repeating and alternating conditions (Figures 6, 7, and 8). In contrast, at the time of the outcome, the neurons exhibited increased activity

with unexpected outcomes only in the repeating condition (Figure 8). This was consistent with the behavioral data: during the overtrained electrophysiological phase, the choices of the mice were stable under the alternating condition, while the choices were updated by past experiences in the repeating condition (Figure 3). Our results suggest the global neural encoding of reward expectations in different behavioral strategies.

The behavioral strategy of repeating a previously rewarded choice, such as the "win-stay lose-switch strategy," has been observed in previous studies with humans,[46] primates,[47] and rodents.[33] We found that mice tended to repeat rewarded choices even when the sound stimuli were randomly presented in the neural condition (Figure 1C). This default strategy was continued to optimize choices in the repeating condition, resulting in the rapid acquisition of proper choice biases. In contrast, mice required several sessions to achieve alternating choice biases in the alternating condition (Figures 1G and S1). A previous study involving a probabilistic foraging task showed that mice used a stimulus-bound model-free strategy in the early phase of training and gradually shifted this strategy to an inference-based strategy.[8] This was consistent with our results in the alternating condition in which the state-based model gradually fit the mouse behavior better than model-free RL, although there was another important possibility that the mice added memory-based states to optimize the choices in alternating condition with model-free strategy (Figure 2H).[34] Similar to our behavioral task, previous studies in free-moving rodents used a perceptual decision-making task with a transition probability in the sensory category[8,33.] These studies trained one animal both in the repeating and alternating conditions and observed the use of a task-specific inference-based strategy in both conditions. In contrast, we trained separate head-fixed mice in either the repeating or alternating condition and found differences in learning speed and strategy between the two task conditions (Figures 1G and 2). Although there were two candidate behavioral strategies in the alternating condition, we verified that mice had different strategies between the repeating and alternating conditions (Figures 2 and 3).

The electrophysiological recordings showed that the choices and outcomes were globally represented in the cortical and subcortical regions, while the sound stimuli were selectively represented in the auditory cortex during sound presentation, irrespective of the task condition (Figures 4E and 4F). We also found that the modulation of neural activity based on reward expectations was globally observed in the cortical and subcortical regions for both the repeating and alternating conditions (Figures 6, 7, and 8). Previous studies have shown that the OFC plays an essential role in both model-free[48] and inference-based flexible behavior.[8] The AC neurons encode task rules and reward expectations.[18,49,50] The HPC, PPC, and STR are involved in either the model-free or model-based strategy.[26,51,52] These studies led to the hypothesis that behavioral strategies are multiplexed and globally encoded in the brain.[15,27,29] However, as many studies have targeted a specific brain region for a specific behavioral strategy, it was unclear whether the neural representations of model-free and inference-based strategies were widely distributed in the brain. Our study suggested that, at least in the overtrained phase, the reward expectations in both the repeating and alternating conditions were globally represented in the brain.

At the outcome timing, in the repeating condition, we found that the neural activity of OFC, PPC, HPC, and AC increased when the mice switched the choice and were rewarded (Figure 8). Switching behavior was uncommon in the repeating condition; thus, reward expectations for switching were lower than those for repetition. These neural representations of unexpected outcomes are important for computing a reward prediction error for value updating in the model-free strategy.[5,53,54] In the alternating condition, we did not observe neural encoding of unexpected outcomes, possibly because of stable mouse behavior (Figure 3B). Additional studies are required to investigate how the inference-based strategy is computed in the brain. Additionally, it is important to investigate neural representation during the learning phase to determine how the different strategies are learned and acquired in the brain.

In summary, we found that mice used the default model-free strategy to bias choice in both the neutral and repeating conditions, while the choice behavior of mice changed to an inference-based strategy in the alternating condition. In the overtrained phase, the neural activity of the frontal and sensory cortices, hippocampus, and striatum was correlated with the reward expectation of both the model-free and inference-based strategies. Neurons in multiple brain regions exhibited increased activity with unexpected outcomes in the repeating condition. These results propose the global encoding of different behavioral strategies in the brain.

Limitations of the study

We did not find clear behavioral evidence of using inference-based strategy in the alternating condition (Figure 3), although the simulated choice behaviors in the hybrid model captured the effects of past events on current choices in the alternating condition (Figure 3H). We found that the AC tended to represent the sound categories of previous trials, which might be required for the inference-based strategy (Figure 5). Further experiments with neural recording and manipulation in sensory cortices with different behavioral tasks are essential to investigate the neural circuit of inference-based strategy. In each session, we found that the behavioral strategies in the repeating and alternating conditions were clearly separated (Figure 3G). However, detailed analyses are required to test whether the behavioral strategy changed trial-by-trial within a session, as reported in previous studies.[14,15] In addition to using computational models to analyze different behavioral strategies in repeating and alternating conditions, further study is required to investigate the neural adaptation to the stimulus sequences in both conditions. Additional experiments with detailed recording of facial and body movements are also important to investigate the neural encoding of different behavioral strategies (Figure S4).

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Akihiro Funamizu (funamizu@iqb.u-tokyo.ac.jp).

Materials availability

This study did not generate new materials.

Data and code availability

- Behavioral and electrophysiology datasets have been deposited at Mendeley data and are publicly available. The DOI is listed in the key resources table.
- All original codes have been deposited at Github and are publicly available as of the date of publication. The DOI is listed in the key resources table.
- Any additional information associated with the data reported in this article is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank Myung Chung and Shun Araki for comments on the article. This work was funded by JSPS Kakenhi (JP21H05243, JP 21H03492, JP 22H04766), AMED JP23wm0525008, and the Uehara Memorial Foundation for A.F.

AUTHOR CONTRIBUTIONS

S.W. collected and analyzed the data and wrote the article. H.G., Y.U., and K.I. analyzed the data. A.F. designed the experiment, analyzed the data, and wrote the article.

DECLARATION OF INTERESTS

The authors declare no conflict of interest.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Surgeries
  - Behavior training
  - Electrophysiological recording and histology
  - Data analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.111182.

Received: February 10, 2024
Revised: September 3, 2024
Accepted: October 14, 2024
Published: October 16, 2024

REFERENCES

1. Funamizu, A. (2021). Integration of sensory evidence and reward expectation in mouse perceptual decision-making task with various sensory uncertainties. iScience *24*, 102826. https://doi.org/10.1016/j.isci.2021.102826.
2. Dayan, P., and Daw, N.D. (2008). Decision theory, reinforcement learning, and the brain. Cognit. Affect Behav. Neurosci. *8*, 429–453. https://doi.org/10.3758/CABN.8.4.429.
3. Sutton, S.R., and Andrew, B.G. (2018). Reinforcement Learning: An Introduction, Second edition.
4. Samejima, K., Ueda, Y., Doya, K., and Kimura, M. (2005). Representation of Action-Specific Reward Values in the Striatum. Science *310*, 1337–1340. https://doi.org/10.1126/science.1115270.
5. Schultz, W., Dayan, P., and Montague, P.R. (1997). A Neural Substrate of Prediction and Reward. Science *275*, 1593–1599. https://doi.org/10.1126/science.275.5306.1593.
6. Daw, N.D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat. Neurosci. *8*, 1704–1711. https://doi.org/10.1038/nn1560.
7. Hampton, A.N., Bossaerts, P., and O'Doherty, J.P. (2006). The Role of the Ventromedial Prefrontal Cortex in Abstract State-Based Inference during Decision Making in Humans. J. Neurosci. *26*, 8360–8367. https://doi.org/10.1523/JNEUROSCI.1010-06.2006.
8. Vertechi, P., Lottem, E., Sarra, D., Godinho, B., Treves, I., Quendera, T., Oude Lohuis, M.N., and Mainen, Z.F. (2020). Inference-Based Decisions in a Hidden State Foraging Task: Differential Contributions of Prefrontal Cortical Areas. Neuron *106*, 166–176.e6. https://doi.org/10.1016/j.neuron.2020.01.017.
9. Akam, T., Rodrigues-Vaz, I., Marcelo, I., Zhang, X., Pereira, M., Oliveira, R.F., Dayan, P., and Costa, R.M. (2021). The Anterior Cingulate Cortex Predicts Future States to Mediate Model-Based Action Selection. Neuron *109*, 149–163.e7. https://doi.org/10.1016/j.neuron.2020.10.013.
10. Dayan, P., and Niv, Y. (2008). Reinforcement learning: The Good, The Bad and The Ugly. Curr. Opin. Neurobiol. *18*, 185–196. https://doi.org/10.1016/j.conb.2008.08.003.
11. Wurm, F., Ernst, B., and Steinhauser, M. (2020). The influence of internal models on feedback-related brain activity. Cognit. Affect Behav. Neurosci. *20*, 1070–1089. https://doi.org/10.3758/s13415-020-00820-6.
12. Pan, X., Sawa, K., Tsuda, I., Tsukada, M., and Sakagami, M. (2008). Reward prediction based on stimulus categorization in primate lateral prefrontal cortex. Nat. Neurosci. *11*, 703–712. https://doi.org/10.1038/nn.2128.
13. Funamizu, A., Kuhn, B., and Doya, K. (2016). Neural substrate of dynamic Bayesian inference in the cerebral cortex. Nat. Neurosci. *19*, 1682–1689. https://doi.org/10.1038/nn.4390.
14. Ashwood, Z.C., Roy, N.A., Stone, I.R., International Brain Laboratory, Urai, A.E., Churchland, A.K., Pouget, A., and Pillow, J.W. (2022). Mice alternate between discrete strategies during perceptual

decision-making. Nat. Neurosci. *25*, 201–212. https://doi.org/10.1038/s41593-021-01007-z.
15. Cazettes, F., Mazzucato, L., Murakami, M., Morais, J.P., Augusto, E., Renart, A., and Mainen, Z.F. (2023). A reservoir of foraging decision variables in the mouse brain. Nat. Neurosci. *26*, 840–849. https://doi.org/10.1038/s41593-023-01305-8.
16. Ito, M., and Doya, K. (2015). Distinct neural representation in the dorsolateral, dorsomedial, and ventral parts of the striatum during fixed- and free-choice tasks. J. Neurosci. *35*, 3499–3514. https://doi.org/10.1523/JNEUROSCI.1962-14.2015.
17. Ito, M., and Doya, K. (2009). Validation of Decision-Making Models and Analysis of Decision Variables in the Rat Basal Ganglia. J. Neurosci. *29*, 9861–9874. https://doi.org/10.1523/JNEUROSCI.6157-08.2009.
18. Guo, L., Weems, J.T., Walker, W.I., Levichev, A., and Jaramillo, S. (2019). Choice-Selective Neurons in the Auditory Cortex and in Its Striatal Target Encode Reward Expectation. J. Neurosci. *39*, 3687–3697. https://doi.org/10.1523/JNEUROSCI.2585-18.2019.
19. Sul, J.H., Jo, S., Lee, D., and Jung, M.W. (2011). Role of rodent secondary motor cortex in value-based action selection. Nat. Neurosci. *14*, 1202–1208. https://doi.org/10.1038/nn.2881.
20. Galea, J.M., Vazquez, A., Pasricha, N., de Xivry, J.J.O., and Celnik, P. (2011). Dissociating the roles of the cerebellum and motor cortex during adaptive learning: The motor cortex retains what the cerebellum learns. Cerebr. Cortex *21*, 1761–1770. https://doi.org/10.1093/cercor/bhq246.
21. Seo, H., and Lee, D. (2009). Behavioral and Neural Changes after Gains and Losses of Conditioned Reinforcers. J. Neurosci. *29*, 3627–3641. https://doi.org/10.1523/JNEUROSCI.4726-08.2009.
22. Doll, B.B., Simon, D.A., and Daw, N.D. (2012). The ubiquity of model-based reinforcement learning. Curr. Opin. Neurobiol. *22*, 1075–1081. https://doi.org/10.1016/j.conb.2012.08.003.
23. Huang, Y., Yaple, Z.A., and Yu, R. (2020). Goal-oriented and habitual decisions: Neural signatures of model-based and model-free learning. Neuroimage *215*, 116834. https://doi.org/10.1016/j.neuroimage.2020.116834.
24. Gläscher, J., Daw, N., Dayan, P., and O'Doherty, J.P. (2010). States versus Rewards: Dissociable Neural Prediction Error Signals Underlying Model-Based and Model-Free Reinforcement Learning. Neuron *66*, 585–595. https://doi.org/10.1016/j.neuron.2010.04.016.
25. Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P., and Dolan, R.J. (2011). Model-Based Influences on Humans' Choices and Striatal Prediction Errors. Neuron *69*, 1204–1215. https://doi.org/10.1016/j.neuron.2011.02.027.
26. Miller, K.J., Botvinick, M.M., and Brody, C.D. (2017). Dorsal hippocampus contributes to model-based planning. Nat. Neurosci. *20*, 1269–1276. https://doi.org/10.1038/nn.4613.
27. Steinmetz, N.A., Zatka-Haas, P., Carandini, M., and Harris, K.D. (2019). Distributed coding of choice, action and engagement across the mouse brain. Nature *576*, 266–273. https://doi.org/10.1038/s41586-019-1787-x.
28. Ishizu, K., Nishimoto, S., Ueoka, Y., and Funamizu, A. (2024). Localized and global representation of prior value, sensory evidence, and choice in male mouse cerebral cortex. Nat. Commun. *15*, 4071. https://doi.org/10.1038/s41467-024-48338-6.
29. Findling, C., Hubert, F., Acerbi, L., Benson, B., Benson, J., Birman, D., Bonacchi, N., Carandini, M., Catarino, J.A., Chapuis, G.A., et al. (2023). Brain-wide representations of prior information in mouse decision-making. https://doi.org/10.1101/2023.07.04.547684.
30. Marbach, F., and Zador, A.M. (2016). A self-initiated two-alternative forced choice paradigm for head-fixed mice. https://doi.org/10.1101/073783.
31. Znamenskiy, P., and Zador, A.M. (2013). Corticostriatal neurons in auditory cortex drive decisions during auditory discrimination. Nature *497*, 482–485. https://doi.org/10.1038/nature12077.
32. Xiong, Q., Znamenskiy, P., and Zador, A.M. (2015). Selective corticostriatal plasticity during acquisition of an auditory discrimination task. Nature *521*, 348–351. https://doi.org/10.1038/nature14225.
33. Hermoso-Mendizabal, A., Hyafil, A., Rueda-Orozco, P.E., Jaramillo, S., Robbe, D., and de la Rocha, J. (2020). Response outcomes gate the impact of expectations on perceptual decisions. Nat. Commun. *11*, 1057. https://doi.org/10.1038/s41467-020-14824-w.
34. Fritsche, M., Majumdar, A., Strickland, L., Liebana Garcia, S., Bogacz, R., and Lak, A. (2024). Temporal regularities shape perceptual decisions and striatal dopamine signals. Nat. Commun. *15*, 7093. https://doi.org/10.1038/s41467-024-51393-8.
35. Bell, A.H., Summerfield, C., Morin, E.L., Malecek, N.J., and Ungerleider, L.G. (2016). Encoding of Stimulus Probability in Macaque Inferior Temporal Cortex. Curr. Biol. *26*, 2280–2290. https://doi.org/10.1016/j.cub.2016.07.007.
36. Funamizu, A., Marbach, F., and Zador, A.M. (2023). Stable sound decoding despite modulated sound representation in the auditory cortex. Curr. Biol. *33*, 4470–4483.e7. https://doi.org/10.1016/j.cub.2023.09.031.
37. Le Merre, P., Esmaeili, V., Charrière, E., Galan, K., Salin, P.A., Petersen, C.C.H., and Crochet, S. (2018). Reward-Based Learning Drives Rapid Sensory Signals in Medial Prefrontal Cortex and Dorsal Hippocampus Necessary for Goal-Directed Behavior. Neuron *97*, 83–91.e5. https://doi.org/10.1016/j.neuron.2017.11.031.
38. MacDonald, C.J., Lepage, K.Q., Eden, U.T., and Eichenbaum, H. (2011). Hippocampal "time cells" bridge the gap in memory for discontiguous events. Neuron *71*, 737–749. https://doi.org/10.1016/j.neuron.2011.07.012.
39. Thomas, M.E., Lane, C.P., Chaudron, Y.M.J., Cisneros-Franco, J.M., and de Villers-Sidani, É. (2020). Modifying the adult rat tonotopic map with sound exposure produces frequency discrimination deficits that are recovered with training. J. Neurosci. *40*, 2259–2268. https://doi.org/10.1523/JNEUROSCI.1445-19.2019.
40. Osako, Y., Ohnuki, T., Tanisumi, Y., Shiotani, K., Manabe, H., Sakurai, Y., and Hirokawa, J. (2021). Contribution of non-sensory neurons in visual cortical areas to visually guided decisions in the rat. Curr. Biol. *31*, 2757–2769.e6. https://doi.org/10.1016/j.cub.2021.03.099.
41. Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C.B., Carandini, M., and Harris, K.D. (2019). Spontaneous behaviors drive multidimensional, brainwide activity. Science *364*, 364. https://doi.org/10.1126/science.aav7893.
42. Zagha, E., Erlich, J.C., Lee, S., Lur, G., O'Connor, D.H., Steinmetz, N.A., Stringer, C., and Yang, H. (2022). The Importance of Accounting for Movement When Relating Neuronal Activity to Sensory and Cognitive Processes. J. Neurosci. *42*, 1375–1382. https://doi.org/10.1523/JNEUROSCI.1919-21.2021.
43. Musall, S., Kaufman, M.T., Juavinett, A.L., Gluf, S., and Churchland, A.K. (2019). Single-trial neural dynamics are dominated by richly varied movements. Nat. Neurosci. *22*, 1677–1686. https://doi.org/10.1038/s41593-019-0502-4.
44. Mathis, A., Mamidanna, P., Cury, K.M., Abe, T., Murthy, V.N., Mathis, M.W., and Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nat. Neurosci. *21*, 1281–1289. https://doi.org/10.1038/s41593-018-0209-y.
45. Liu, Y., Xin, Y., and Xu, N.l. (2021). A cortical circuit mechanism for structural knowledge-based flexible sensorimotor decision-making. Neuron *109*, 2009–2024.e6. https://doi.org/10.1016/j.neuron.2021.04.014.
46. Bonawitz, E., Denison, S., Gopnik, A., and Griffiths, T.L. (2014). Win-Stay, Lose-Sample: A simple sequential algorithm for approximating Bayesian inference. Cognit. Psychol. *74*, 35–65. https://doi.org/10.1016/j.cogpsych.2014.06.003.
47. Schusterman, R.J. (1963). The use of strategies in 2-choice behavior of children and chimpanzees. J. Comp. Physiol. Psychol. *56*, 96–100.
48. Jones, J.L., Esber, G.R., McDannald, M.A., Gruber, A.J., Hernandez, A., Mirenzi, A., and Schoenbaum, G. (2012). Orbitofrontal Cortex Supports Behavior and Learning Using Inferred But Not Cached Values. Science *338*, 953–956. https://doi.org/10.1126/science.1227489.
49. Francis, N.A., Winkowski, D.E., Sheikhattar, A., Armengol, K., Babadi, B., and Kanold, P.O. (2018). Small Networks Encode Decision-Making in Primary Auditory Cortex. Neuron *97*, 885–897.e6. https://doi.org/10.1016/j.neuron.2018.01.019.
50. Zempeltzi, M.M., Kisse, M., Brunk, M.G.K., Glemser, C., Aksit, S., Deane, K.E., Maurya, S., Schneider, L., Ohl, F.W., Deliano, M., and Happel, M.F.K. (2020). Task rule and choice are reflected by layer-specific processing in rodent auditory cortical microcircuits. Commun. Biol. *3*, 345. https://doi.org/10.1038/s42003-020-1073-3.
51. McDannald, M.A., Lucantonio, F., Burke, K.A., Niv, Y., and Schoenbaum, G. (2011). Ventral Striatum and Orbitofrontal Cortex Are Both Required for Model-Based, But Not Model-Free, Reinforcement Learning. J. Neurosci. *31*, 2700–2705. https://doi.org/10.1523/JNEUROSCI.5499-10.2011.
52. Zhong, L., Zhang, Y., Duan, C.A., Deng, J., Pan, J., and Xu, N.L. (2019). Causal contributions of parietal cortex to perceptual decision-making during stimulus categorization. Nat. Neurosci. *22*, 963–973. https://doi.org/10.1038/s41593-019-0383-6.
53. Schultz, W. (2016). Dopamine reward prediction error coding. Dialogues Clin. Neurosci. *18*, 23–32. https://doi.org/10.31887/DCNS.2016.18.1/wschultz.
54. Sambrook, T.D., Hardwick, B., Wills, A.J., and Goslin, J. (2018). Model-free and model-based reward prediction errors in EEG.

Neuroimage *178*, 162–171. https://doi.org/10.1016/j.neuroimage.2018.05.023.

55. Fiáth, R., Márton, A.L., Mátyás, F., Pinke, D., Márton, G., Tóth, K., and Ulbert, I. (2019). Slow insertion of silicon probes improves the quality of acute neuronal recordings. Sci. Rep. *9*, 111. https://doi.org/10.1038/s41598-018-36816-z.
56. Masset, P., Ott, T., Lak, A., Hirokawa, J., and Kepecs, A. (2020). Behavior- and Modality-General Representation of Confidence in Orbitofrontal Cortex. Cell *182*, 112–126.e18. https://doi.org/10.1016/j.cell.2020.05.022.
57. Lak, A., Okun, M., Moss, M.M., Gurnani, H., Farrell, K., Wells, M.J., Reddy, C.B., Kepecs, A., Harris, K.D., and Carandini, M. (2020). Dopaminergic and Prefrontal Basis of Learning from Sensory Confidence and Reward Value. Neuron *105*, 700–711.e6. https://doi.org/10.1016/j.neuron.2019.11.018.
58. Funamizu, A., Ito, M., Doya, K., Kanzaki, R., and Takahashi, H. (2012). Uncertainty in action-value estimation affects both action choice and learning rate of the choice behaviors of rats. Eur. J. Neurosci. *35*, 1180–1189. https://doi.org/10.1111/j.1460-9568.2012.08025.x.
59. Hattori, R., Danskin, B., Babic, Z., Mlynaryk, N., and Komiyama, T. (2019). Area-Specificity and Plasticity of History-Dependent Value Coding During Learning. Cell *177*, 1858–1872.e15. https://doi.org/10.1016/j.cell.2019.04.027.
60. Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. (2011). Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. J. Stat. Software *39*, 1–13. https://doi.org/10.18637/jss.v039.i05.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Behavioral and electrophysiology data	This paper, Mendeley Data	https://doi.org/10.17632/vf4b4bmzjp.1
Experimental models: Organisms/strains		
CBA/J mouse	Jackson Laboratory	Jax stock 000656; RRID: IMSR_JAX: 000656
Software and algorithms		
MATLAB 2022b	Mathworks	https://jp.mathworks.com
Code	This paper, Github	https://github.com/funamizu-lab/Wang_et_al_2024.git
KiloSort-3	Cortex lab	https://github.com/cortex-lab/KiloSort
Phy	Cortex lab	https://github.com/cortex-lab/phy
Affinity Designer 2.1	Serif	https://affinity.serif.com
Other		
Bpod framework (control for behavioral task)	Sanworks	r0.5
Microphone for sound calibration	Brüel and Kjaer	Type 4939
Speaker	Avisoft Bioacoustics	#60108

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

All animal procedures were approved by the Animal Care and Use Committee at the Institute for Quantitative Biosciences (IQB), the University of Tokyo. Mice were housed in a temperature-controlled room with a 12 h/12 h light/dark cycle. All the experiments were performed during the dark cycle.

Male CBA/J mice (n = 26, Strain #000656; The Jackson Laboratory), aged 8 to 15 weeks at the start of behavioral training, were used for the experiments. 26 mice were allocated for behavioral experiments. 14 out of the 26 mice were subjected to electrophysiological recording after the behavioral experiments. Before surgery, 3 mice were housed in one cage. Mice were allowed *ad libitum* access to food, while water intake was restricted to 1.5 mL per day. On weekends, the mice were given 3 mL of extra water and free access to 1.5% citric acid water to prevent dehydration. Mice were caged in isolation after craniotomy.

METHOD DETAILS

Surgeries

The surgical procedures were described in our previous research.[1,28] In summary, the surgery had two steps. First, a custom-designed head bar was implanted for behavioral training. Second, a craniotomy was performed for electrophysiological recording.

For head bar implantation, the mice were anesthetized via intraperitoneal injection of a mixture of medetomidine (0.3 mg/kg), midazolam (4.0 mg/kg), and butorphanol (5.0 mg/kg). Meloxicam (2.5 mg/kg) and eye ointment were also used. Mice were placed in a stereotaxic apparatus. The scalp was removed above the entire cortical area. We cleaned the skull with povidone iodine and hydrogen peroxide. We attached the head bar to the skull with Superbond adhesive (Sun Medical or Parkell S380) and cyanoacrylate glue (Zap-A-Gap, PT03).[43]

For craniotomy, the mice were anesthetized with isoflurane (2% for induction, 1.5% for maintenance). We targeted 6 brain regions: the OFC, STR, M1, PPC, HPC, and AC. The OFC sites were +2.6 mm anterior-posterior (AP) and ±1.4 mm medio-lateral (ML) from bregma. The STR and M1 sites were +0.8 mm AP and ±1.5 mm ML. The PPC and HPC sites were −2.0 mm AP and ±1.7 mm ML. The AC sites were −3.0 mm AP and ±3.8 mm ML (Figure S2). We drilled a small hole (0.5–0.8 mm in diameter) through cyanoacrylate glue and the skull to expose the brain surface and removed the dura mater. We covered the brain surface with agar dissolved in PBS followed by silicone oil and Kwik-Sil (World Precision Instruments) to prevent drying.

Behavior training

Behavioral apparatus

We performed behavioral experiments inside a custom-made training box or a sound-attenuating booth (O'hara, Inc.). After recovering from the head bar implantation, the mice were head-fixed and placed on a custom cylinder treadmill. We presented sound stimuli pre-calibrated with a Brüel and Kjaer microphone (Type 4939) from an Avisoft Bioacoustics speaker (#60108) positioned to the right front of the mice. Two

spouts were placed in front of the mice to deliver sucrose water. The licking behavior of the mice was detected using electrical or infrared sensors.[30] We used custom-made MATLAB (MathWorks) programs on Bpod r0.5 (https://sanworks.io) on Windows OS.

Tone frequency discrimination task with probabilistic alternation of tone category with transition probabilities

Like in our previous studies,[1,28] each trial began by retracting the two spouts away from the mice. After a random interval of 1–2 s, a sound stimulus in the form of a tone cloud was presented.[30–32] The intensity of the tone cloud in each trial remained constant but was sampled from 60, 65, or 70 dB SPL (the sound pressure level in decibels with respect to 20 μPa). The duration of the tone cloud was held constant at 0.6 s. The tone cloud was a mixture of low-frequency (5–10 kHz) and high-frequency (20–40 kHz) tones.

After the sound ended, the two spouts were immediately displayed to the mice. Mice selected either the left or right spout depending on the dominant tone frequency. The association between the tone category and rewarded choice was determined for each mouse. A correct choice provided 2.4 μL of 10% sucrose water. An incorrect choice triggered a 0.2-s noise burst from the speaker. If the mouse failed to select a spout within 15 s, a new trial started.

Our behavioral task had 4 steps.

1) The initial step involved training the mice to discriminate between the 90% low-frequency and 90% high-frequency tone clouds. We trained 26 mice in the initial step.
2) After the mice were able to discriminate the sounds, the neutral condition started. We used 6 tone clouds (0, 20, 35, 65, 80, and 100% high-frequency tones) with presentation probabilities of 25%, 12.5%, 12.5%, 12.5%, 12.5%, and 25%, respectively. The neutral condition had a transition probability 'p', which controlled how often the tone category of the previous trial alternated in the current trial. We set '$p = 0.5$' to randomly present the tone category in each trial. 7 out of the 26 mice skipped step 2 and directly completed step 3.
3) After the mouse experienced at least one training session in the neutral condition or when the percentages of correct responses for tone clouds that were 90% high tones and 90% low tones were both greater than 80% in the initial step, we assigned the mouse to either the repeating or alternating condition (22 out of 26 mice). In both conditions, we used 6 tone clouds (0, 25, 45, 55, 75, and 100% high-frequency tones) with presentation probabilities of 25%, 12.5%, 12.5%, 12.5%, 12.5%, and 25%, respectively. The repeating condition had a transition probability of '$p = 0.2$', in which the same tone category was frequently presented. In contrast, in the alternating condition, we set the transition probability as '$p = 0.9$', where the tone category was alternated every trial in 90% of trials. In both the repeating and alternating conditions, each session started with 40 trials of only 100% low- or high-tone clouds. The repeating condition switched the tone category in every 10 trials, while the alternating condition switched the category every trial. After the first 40 trials, the above conditions for the 6 tone clouds and transition probabilities started. 3 mice did not show transition probability-dependent choice biases, as analyzed with a psychometric function (STAR Methods, Behavioral analysis),[1] in the alternating condition after more than 11 sessions and were not used for the analyses.
4) After the mice experienced at least 9 sessions of either the repeating or alternating condition, we started the prerecording step. This step gradually increased the interval between the end of the sound and the spout approaching until the interval reached 0.5 s. The transition probability of the alternating condition was set to '$p = 0.8$'. 14 mice completed the prerecording step.

Electrophysiological recording and histology

The electrophysiological recording performed with Neuropixels 1.0 (IMEC), and the histological analysis were described in detail in our previous study.[28] In summary, we inserted the Neuropixels probe 1 to 3 times in each hole in both the left and right hemispheres of the brain. For the OFC recordings, the probe was tilted 5° in the medial direction. For the AC recordings, the probe was tilted 18° in the lateral direction. The angle was 0° for the PPC, HPC, STR, and M1 recordings. To identify the probe location in the post hoc fixed brain, the probe was soaked in a diluted solution of CM-DiA or DiI (Thermo Fisher Product #D3883 or #V22888). The probe was manually lowered to the brain surface through a mixture of agar and PBS at a speed of 120 μm/min (MPC-200 Controller and ROE, Sutter Instrument).[55] The brain surface was defined based on the depth at which spikes were initially observed at the recording electrodes. The 384 electrodes from the tip of the Neuropixels probe were used for recording. The Open-Ephys GUI acquired the neural data at a sampling rate of 30 kHz with a gain of 500 (PXIe acquisition module, IMEC). The task events, including treadmill rotations, were sampled at 2.5 kHz (BNC-2110, National Instruments). After the recording was complete, the probe was slowly extracted, and the hole was covered with Kwik-Sil.

After the electrophysiological recording, the mice were deeply anesthetized with isoflurane (5%) and further anesthetized with a mixture of 1.5 mg/kg medetomidine, 20 mg/kg midazolam, and 25 mg/kg butorphanol. Mice were perfused with 10% formalin solution. Brain sections were sliced with a vibratome to a thickness of 100 μm (VT1000S; Leica Biosystems) and mounted with DAPI mounting medium (Vector Laboratories, Cat. No. H-1200). The probe locations were captured with a confocal laser scanning microscope (FV3000, Olympus) at 4× magnification (Figure 4).

Data analysis

Number of mice and sessions in behavioral tasks

We used the sessions for analyses for which (i) the correct response rate of the mouse for both the 100% low and 100% high tones exceeded 75% and (ii) the total reward amount in one session was above 600 μL. Under neutral conditions, we analyzed 28 sessions from 16 out of

19 mice (Figure 1); the sessions in 3 mice did not exceed the accuracy rate of 75%. Under the repeating and alternating conditions, we analyzed 113 and 141 sessions, respectively, from 8 to 11 mice (Figure 1). The behavioral results of (i) all the sessions (neutral condition: 35 sessions in 19 mice; repeating and alternating conditions: 139 and 165 sessions from 8 to 11 mice) and (ii) the sessions with correct rates over 80% (108 and 117 sessions from 8 to 10 mice) are shown in Figure S1.

Behavioral analysis

We analyzed the choice behavior of mice with a psychometric function based on our previous study.[1] The psychometric function models the perceptual uncertainty of mice with a truncated Gaussian ranging between 0 and 1.[56] Our model investigated whether the choice biases of mice depended on the rewarded side in the previous trial. We tested whether the psychometric function with a choice-bias parameter better fit the mouse choices than that without the parameter ($p < 0.01$ in the likelihood ratio test). We investigated when the mouse started to bias the choices in the repeating and alternating conditions (Figure 1G).

We also investigated how the events in past trials affected the choice in current trial with a logistic regression (MATLAB: fitglm) (Figures 1E, 1I, and 3C):[34]

$$P(right,t) = \frac{1}{1+e^{-y(t)}}$$

$$y(t) = \beta_0 + \beta_1 E_{right}(t) + \sum_{i=1}^{n}\{\beta_{2,i}C(t-i) + \beta_{3,i}I(t-i)\} \quad \text{(Equation 1)}$$

$P(right,t)$ was the probability of choosing right at trial t. $E_{right}(t)$ was the proportion of tone frequency associated with a rightward choice in a tone cloud. $C(t-i)$ and $I(t-i)$ represent the correct and incorrect choices at the trial $t-i$. C was -1 and 1 for the previous correct left- and right-choices, respectively, while 0 for the incorrect choices. I was -1 and 1 for the previous incorrect left- and right-choices, while 0 for the correct choices. $\beta_0, \beta_1, \beta_{2,i}, \beta_{3,i}$ were the regression coefficients. We analyzed the likelihood in each session L with Equations 14 and 15 (STAR Methods, Model Comparison). We used the likelihood ratio test to quantify whether the additional past trials from 1- to 5-back (i.e., from $n = 1$ to $n = 5$) improved the choice prediction (Figures 1H and S1I).[28] We also investigated the regression coefficients of model with 5-back trials (Figure 1I).

Behavioral model

The behavioral models were based on signal detection theory (SDT). The behavioral task required the mice to estimate the hidden state (S) of left-rewarded ($S = L$) or right-rewarded ($S = R$) based on the sensory evidence of the tone cloud. SDT shows that both the expected outcome in each state and the belief state probability are essential for optimizing choices.[1,2] The following model-free reinforcement learning (RL) model and state-based model estimated the expected outcome and state probability, respectively.

Model-free reinforcement learning (RL) model. The RL model updated the expected outcome of left and right choice in each state $Q_{a,S}$, defined as the prior value,[1,28,57] while the belief state probability was fixed. We denoted the choice as a. We assumed that there were only left-rewarded and right-rewarded states; the prior values satisfied the criteria $Q_{left,R} = Q_{right,L} = 0$. We simplified $Q_{left,L}$ and $Q_{right,R}$ as Q_{left} and Q_{right}, respectively. The model used prior values to compute the decision threshold x_0 in each trial (t) with a softmax function and an inverse temperature parameter β:

$$x_0(t) = \frac{exp(\beta Q_{left}(t))}{exp(\beta Q_{left}(t)) + exp(\beta Q_{right}(t))} \quad \text{(Equation 2)}$$

The softmax equation modeled a perceived reward size that might be different from the actual amount of water. The right-choice probability at trial t, $P(right,t)$, was estimated from a perceptual uncertainty σ and a bias parameter d:

$$P(right,t) = \int_{x_0(t)+d}^{1} ZN\left(x \middle| E_{right}(t), \sigma^2\right) dx \quad \text{(Equation 3)}$$

$E_{right}(t)$ was the proportion of tone frequency associated with a rightward choice in a tone cloud. Z truncated the Gaussian distribution between 0 and 1, here and hereafter. We updated the prior value with forgetting Q-learning[17]:

$$Q_a(t+1) = \begin{cases} Q_a(t) + \alpha(r(t) - Q_a(t)) & if\ a = a(t) \\ (1-\alpha)Q_a(t) & if\ a \neq a(t) \end{cases} \quad \text{(Equation 4)}$$

where α was the learning rate. $r(t)$ was the outcome at trial t. The initial prior value for each choice was the amount of reward (i.e., 2.4).

State-based model. The state-based model had the belief of state probability $P(S)$ in each trial by estimating and updating the transition of state $P_{transition}$ in every trial.[7] The prior values were fixed. Bayesian inference provided the decision threshold of choice based on $P(S)$. First, the likelihood of a sensory stimulus x in state S_i, $P(x|S_i)$, was defined as follows:

$$P(x|S_i) = \int_0^1 P(E_j|S_i) ZN(x|E_j, \sigma^2) dE_j \quad \text{(Equation 5)}$$

$P(E_j|S_i)$ was the probability of tone cloud E_j in a given state S_i. σ was a free parameter for perceptual uncertainty. The posterior probability $P(S_i|x)$ was calculated with the Bayes rule:

$$P(S_i|x) \propto P(x|S_i) P(S_i) \quad \text{(Equation 6)}$$

The decision threshold x_0 satisfied $P(S_L|x_0) = P(S_R|x_0)$. We used a softmax equation to add flexibility to the threshold x_0 with an inverse temperature parameter β:

$$x_0 = \frac{x_0}{exp(\beta x_0) + exp(\beta(1 - x_0))} \quad \text{(Equation 7)}$$

Based on x_0 and the bias choice parameter d, the model estimated the choice in each trial with Equation 3.

After the model received the outcome at trial t, the state transition $P_{transiton}$ was updated based on the true state of trial t and t-1:

$$P_{transition}(t+1) = \begin{cases} P_{transition}(t) + \alpha(1 - P_{transition}(t)) & if\ S_i(t) \neq S_i(t-1) \\ P_{transition}(t) + \alpha(0 - P_{transition}(t)) & if\ S_i(t) = S_i(t-1) \end{cases} \quad \text{(Equation 8)}$$

where α was the learning rate. The model computed the belief state of trial t+1 based on the transition probability and the true state at t. The initial prior belief of left- and right-rewarded states was 0.5. The initial transition probability was 0.5 for the learning phase of the repeating and alternating conditions, while it was the true transition probability for the overtrained phase.

Hybrid model. The hybrid model computed the hybrid value $Hybrid_a$ by combining the prior value Q_a in the RL model and the belief of state probability $P(S)$ at trial t.

$$Hybrid_{left,t} = (1 - P_{mix}) Q_{left,t} \Big/ \left(Q_{left,t} + Q_{right,t}\right) + P_{mix} P(L)_t$$

$$Hybrid_{right,t} = (1 - P_{mix}) Q_{right,t} \Big/ \left(Q_{left,t} + Q_{right,t}\right) + P_{mix} P(R)_t \quad \text{(Equation 9)}$$

P_{mix} was the mixed ratio of state probability in the hybrid value.

The hybrid model used the hybrid value, instead of Q_a, to compute the decision threshold x_0 with the softmax function of Equation 2 and estimated the right-choice probability with Equation 3. The prior value was updated with forgetting Q-learning with Equation 4, while the transition probability was updated with Equation 8 for the belief-state computation.

Memory-based model-free RL model (memory model and forgetting memory model). We modeled the mice choices with a memory-based model-free RL model (memory model) based on a previous study.[34] The memory model computed the decision threshold x_0 and the right-choice probability at trial t, $P(right,t)$, in the same way as the RL model did (Equations 2 and 3). However, the prior values depended on the current choice and the memory of the choice in previous trial[34]:

$$Q_a = q_a + \sum_{j \in \{left,right\}} M_j \cdot q_{a,M_j} \quad \text{(Equation 10)}$$

where q_a was the prior value based on the current choice a. q_{a,M_j} was the value based on the choice in the current (a) and the previous trial (M_j). M_{left} and M_{right} were the memory strength for the previous correct trial:

$$M_{left} = \begin{cases} \lambda & if\ prev.\ choice\ was\ left\ and\ rewarded \\ 0 & otherwise \end{cases}$$

$$M_{right} = \begin{cases} \lambda & if\ prev.\ choice\ was\ right\ and\ rewarded \\ 0 & otherwise \end{cases} \quad \text{(Equation 11)}$$

λ was bounded between 0 and 1 (no use and perfect knowledge of the previous rewarded choice). We updated the values with a standard value updating in the RL model[34]:

$$q_a(t+1) = \begin{cases} q_a(t) + \alpha(r(t) - Q_a(t)) & if\ a = a(t) \\ q_a(t) & if\ a \neq a(t) \end{cases}$$

$$q_{a,M_{left}}(t+1) = \begin{cases} q_{a,M_{left}}(t) + \alpha \cdot M_{left} \cdot (r(t) - Q_a(t)) & if\ a = a(t) \\ q_{a,M_{left}}(t) & if\ a \neq a(t) \end{cases} \quad \text{(Equation 12)}$$

$$q_{a,M_{right}}(t+1) = \begin{cases} q_{a,M_{right}}(t) + \alpha \cdot M_{right} \cdot (r(t) - Q_a(t)) & if\ a = a(t) \\ q_{a,M_{right}}(t) & if\ a \neq a(t) \end{cases}$$

We quantified that our version of memory model better fit to the mice choices in our study compared to the original model in the previous study[34] (average Bayesian information criterion (BIC) of modified and original models: repeating condition with 51 sessions, 369.7 and 436.1, p = 9.1e-9 in the Wilcoxon signed rank test; alternating condition with 127 sessions, 477.0 and 530.8, p = 2.5e-22).

We also developed a forgetting-memory model (f-memory model) by updating the values with forgetting Q-learning instead of Equation 12:

$$q_a(t+1) = \begin{cases} q_a(t) + \alpha(r(t) - Q_a(t)) & if\ a = a(t) \\ (1-\alpha) q_a(t) & if\ a \neq a(t) \end{cases}$$

$$q_{a,M_{left}}(t+1) = \begin{cases} q_{a,M_{left}}(t) + \alpha \cdot M_{left} \cdot (r(t) - Q_a(t)) & if\ a = a(t) \\ (1-\alpha) q_{a,M_{left}}(t) & if\ a \neq a(t) \end{cases} \quad \text{(Equation 13)}$$

$$q_{a,M_{right}}(t+1) = \begin{cases} q_{a,M_{right}}(t) + \alpha \cdot M_{right} \cdot (r(t) - Q_a(t)) & if\ a = a(t) \\ (1-\alpha) q_{a,M_{right}}(t) & if\ a \neq a(t) \end{cases}$$

Previous studies show that the choice behavior of rodents fit to a forgetting Q-learning than a standard Q-learning model.[16,58,59]

Model comparison

We defined the likelihood $l(t)$ from the estimated choice probability $P(right, t)$ in each trial:

$$l(t) = \begin{cases} P(right, t) & if\ a(t) = right \\ 1 - P(right, t) & if\ a(t) = left \end{cases} \quad \text{(Equation 14)}$$

We then analyzed the likelihood in each session L using the trials without the first 40 trials in each session:

$$L = \prod_{t=1}^{T} l(t) \quad \text{(Equation 15)}$$

where T was the number of trials. The model parameters were fit to achieve the maximum likelihood. We first used the Bayesian information criterion (BIC) to identify the necessary parameters in the RL, state-based, hybrid, memory, or f-memory model. We also used the BIC to compare the performance across models (Figure 2):

$$BIC = -2\log(L) + k\log(T) \quad \text{(Equation 16)}$$

where k was the number of free parameters.

Number of mice and sessions of electrophysiological neural recording

Same as the behavioral data analyses, we used the sessions for analyses when (i) the percentage of correct responses for both the 100% low tones and 100% high tones stimuli were greater than 75% and (ii) the total reward amount in one session was at least 600 μL. We analyzed OFC neurons from 17 to 30 sessions from 5 to 7 mice in the repeating and alternating conditions, respectively; PPC and HPC neurons from 18 to 39 sessions from 4 to 9 mice; and AC neurons from 16 to 31 sessions from 4 to 8 mice. The STR and M1 neurons from 39 sessions from 7 mice were analyzed only in the alternating condition. 13 and 36 sessions in the repeating and alternating conditions, respectively, were excluded from the analyses.

Electrophysiology data analysis

Spike sorting and manual curation were performed with KiloSort-3 on MATLAB (https://github.com/cortex-lab/KiloSort) and Phy on Python (https://github.com/cortex-lab/phy). KiloSort-3 spike sorting tracked the approximate depth of spikes from each unit during a session. We defined the depth of each unit from the approximate location of the probe and the electrode position, which measured the maximum amplitude of spikes on average.

For a Neuropixels probe for the OFC, we used the units for analyses when the estimated spike depth from the brain surface was less than 1.9 mm. For the PPC recording probe, we used units for analyses when the estimated spike depth was less than 1.0 mm. For the STR recording probe, we analyzed the units when the estimated spike depth was greater than 1.5 mm. For the M1 probe, we used units when the estimated

spike depth was less than 1.5 mm. For the HPC, we used the units for analyses when the estimated spike depth was between 1 and 2.3 mm. We analyzed all the units recorded by the AC probe.

We identified task-relevant neurons that exhibited increased activity during the task ($p < 1.0e\text{-}10$ in the one-sided Wilcoxon signed rank test) compared to baseline activity. For sound-aligned activity, task-relevant neurons exhibited increased activity in at least one time window (0.1 s) between −1.5 s and 2.5 s from sound onset (40 windows). For choice-aligned activity, task-relevant neurons exhibited increased activity in at least one time window between −0.5 and 2.5 s from the choice (30 windows, in total 40 + 30 = 70 windows). The baseline for the sound-aligned activity was −0.2 to 0 s from spout removal, i.e., before trial initiation. The baseline for the choice-aligned activity was −0.2 to 0 s from the time the spout approached, i.e., between the end of the sound and making a choice.

Among the task-relevant neurons, we focused on the neurons that exhibited significantly increased activity (1) between −0.6 and 0 s from sound onset (before sound), (2) between 0 and 0.6 s from sound onset (during sound), and (3) between 0 and 1.0 s from choice (during outcome).

We analyzed the choice index of each neuron based on the left- and right-choice trials:

$$Choice\ Index = \frac{mean(spike\ (right\ choice\ trials)) - mean(spike\ (left\ choice\ trials))}{mean(spike\ (right\ choice\ trials)) + mean(spike\ (left\ choice\ trials))} \quad \text{(Equation 17)}$$

The choice index ranged between −1 and 1 and was independently analyzed for correct and incorrect trials (Figures 7 and 8). In the analyses of neural activity before sound presentation (Figure 5C), we analyzed the choice indices of the previous and current trials to determine whether the neurons represented previous or current choices: with respect to the choice indices in previous correct trials, we independently analyzed the choice indices in the current left- or right-choice trials and averaged the values (previous choice indices). For the current choice index, we independently analyzed the choice index in the previous left or right correct choice trials and averaged the values.

We defined the preferred side of the task-relevant neurons based on (i) the choice index in the correct trials and (ii) the activity difference between the left and right correct choice trials. For the neural activity before sound, we defined the preferences of neurons based on the choice index in previous choice (Figures 5 and 6). For the neural activity during sound and outcome, the preference was defined based on the choice index in current choice (Figures 7 and 8). Second, the activity on the preferred side was greater than that on the nonpreferred side (Mann−Whitney U test, $p < 0.01$). For example, during sound, when the choice index of a neuron was less than 0 and the activity in the correct left-choice trials was greater than that in the correct right-choice trials, the preferred side was defined as the left side. If neurons did not show a significant difference in activity between the left and right correct choice trials ($p > 0.01$), we defined them as non-side-preferred neurons (Figures 5, 6, 7, 8, S5–S8).

Facial motion analysis

During the electrophysiological neural recording, we recorded the facial movements of mice in 11 and 77 sessions from 1 to 5 mice in the repeating and alternating conditions, respectively, out of 51 and 127 sessions from 5 to 9 mice. The sessions in the repeating condition were excluded from the regression analyses in Figures 6E, 7H, and 8H due to the limited number of mice and sessions. In the alternating condition, we analyzed the OFC neurons in 18 sessions from 4 mice; the PPC and HPC neurons in 22 sessions from 5 mice; the AC neurons in 22 sessions from 5 mice; the STR and M1 neurons in 15 sessions from 4 mice.

We captured the facial movements of mice with one camera with a sampling rate of 140 Hz. We extracted the 9 facial features (left/right eyes, left/right whiskers, root/tip of tongue, right/left/tip of nose) and the spouts movement by DeepLabCut (DLC).[44] The spouts movement was used to temporally align the movie and task parameters in 55 sessions. In the other 33 sessions, the movie captured a 660-nm LED stimulus at the trial start for the temporal alignment. In each of 9 facial features, we first detected the frame-by-frame changes in the xy coordinates. The velocity of each facial feature was defined as the square root of the sum of the squared changes. We defined the facial motion strengths as the sum of velocities in all the 9 facial features, and standardized the strengths between 0 and 1 in each session.

We investigated whether the facial motion strength F_t in a specific time window at trial t correlated to the previous and current choices (C), sounds (S), and outcomes (O) (Figure S4C):

$$F_t = \beta_0 + \beta_1 C_t + \beta_2 S_t + \beta_3 O_t + \beta_4 C_{t-1} + \beta_5 S_{t-1} + \beta_6 O_{t-1} \quad \text{(Equation 18)}$$

β_{0-6} were regression coefficients.

Regression analysis

A generalized linear model (GLM) was used to analyze whether the activity of a specific time window in the task represented the choice (C), sound (S), or outcome (O) in both the current and previous trials, in addition to the running speed (R) (MATLAB, glmfit with Poisson distribution, $p < 0.01$) (Figure 4C):

$$Spike_t = \exp(\beta_0 + \beta_1 C_t + \beta_2 S_t + \beta_3 O_t + \beta_4 C_{t-1} + \beta_5 S_{t-1} + \beta_6 O_{t-1} + \beta_7 R_t) \quad \text{(Equation 19)}$$

$Spike_t$ was the number of spikes at trial t in the time windows of either before sound presentation, during the sound, or during the outcome. β_{0-7} were the regression coefficients. We used the MATLAB software package glmnet with Poisson distribution and L1 regularization (https://glmnet.stanford.edu/index.html).[60] The deviance of GLM was validated with 10-fold cross validation (CV). The CV was repeated

100 times and investigated the average deviance to reduce noise from random grouping of trials, defined as the GLM performance. To test whether the GLM analysis captured the neural encoding of task variables, we shuffled the neural activity in each trial and performed the GLM analysis. The CV with shuffled activity was repeated 200 times. We analyzed the proportion of task-relevant neurons in which the GLM performance was lower than the distribution of deviances in the 200 CVs in shuffled activity ($p < 0.025$) (Figure 4D).

We then analyzed whether the task-relevant neurons represented each task variable. In the regression analysis of Equation 19, we removed one of the six task variables (i.e., C_t, S_t, O_t, C_{t-1}, S_{t-1}, O_{t-1}) and analyzed the deviance of GLM with 10-fold CV.[40] The CV was repeated 100 times to investigate the mean and standard deviation of 100 deviances. When the GLM performance of full parameters was lower than the mean - 1.96 × standard deviation of deviances in the one-parameter-removed GLM, we defined that the neuron represented the removed task variable (Figures 4E and 4F).

In the sessions with face capturing (i.e., 11 and 77 sessions in the repeating and alternating conditions), we additionally investigated whether the neurons represented the task variables and the facial motion strengths (Figure S4D):

$$Spike_t = \exp\left(\beta_0 + \beta_1 C_t + \beta_2 S_t + \beta_3 O_t + \beta_4 C_{t-1} + \beta_5 S_{t-1} + \beta_6 O_{t-1} + \beta_7 R_t + \beta_8 F_t\right) \quad \text{(Equation 20)}$$

The analyses were same as for Equation 19.

As the facial movements affected the neural encoding (Figure S4D), we used a regression analysis and investigated whether the choice-sequence dependent modulations of neural activity were affected by the facial motion strengths (MATLAB, glmfit with Poisson distribution) (Figures 6E, 7H, and 8H):

$$Spike_t = \exp\left(\beta_0 + \beta_1 F_t + \sum_{a \in \{left,right\}} \sum_{j \in \{left,right\}} \beta_{a,M_j} \cdot C_{a,M_j,t}\right) \quad \text{(Equation 21)}$$

$C_{a,M_j,t}$ was the combination of previous (M_j) and current (a) choices at trial t. For example, when a mouse selected the left and left for the previous and current choices, $C_{left,left,t}$ was 1 and the other 3 $C_{a,M_j,t}$ were 0. We investigated the regression coefficients for the repeated and switched choices for the preferred and nonpreferred sides in each neuron.

QUANTIFICATION AND STATISTICAL ANALYSIS

We used MATLAB 2022b for all the analyses except for spike sorting, which was performed with KiloSort 3 on Python. The statistical details are shown in the Results section, the figures, and the figure legends. Solid lines and shaded areas are represented as means ± standard deviations (SD) or standard errors (SEM), respectively. In the analyses of psychometric function, we used the likelihood ratio test to investigate whether the additional parameter of choice bias significantly improved the fit to mouse choices. Model fitting of the RL, state-based, hybrid, memory and f-memory models were performed with the Bayesian information criterion (BIC). For the behavioral analyses of multiple sessions in each mouse, we employed the linear mixed-effects model (MATLAB: fitlme). In other analyses, we used two-sided nonparametric statistical tests. We used the MATLAB glmfit function to analyze the proportion of neurons representing each task variable (Figure 4). We used the chi-square test to compare the proportion of neurons representing task variables.