The Periaqueductal Gray Selectively Supports Reversal Learning During a Flexible Discrimination Task in Mice

Daniela Lichtman,[1,2] Eyal Bergmann,[1,2] Jonathan Nicholas,[3] Raphael T. Gerraty,[2] and Itamar Kahn[2,*]

[1]Department of Neuroscience, Rappaport Faculty of Medicine, Technion – Israel Institute of Technology, Haifa 31096, Israel
[2]Department of Neuroscience, Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY
[3]Department of Psychology, New York University, New York, NY

*Corresponding Author:
ik2508@columbia.edu
Mortimer B. Zuckerman Mind Brain Behavior Institute
Columbia University
3227 Broadway St., New York, NY 1002

Abstract

Flexible, goal-directed behavior depends on the ability to update value representations in response to changing contingencies. This ability depends on distributed brain networks, calling for use of whole-brain imaging. Widely used in human research, the challenge of functional fMRI in rodents has only recently been properly addressed, opening the door to whole-brain imaging of behaving mice. Here, we used functional MRI in mice performing a go/no-go odor discrimination task, we compared neural activity during initial cue-reward learning (acquisition) and subsequent contingency reversal. Trial-by-trial estimates of state-action values from a model-free reinforcement-learning algorithm allowed us to dissociate acquisition from reversal-related signals, revealing that ventral striatal responses tracked expected value during acquisition, whereas reversal learning additionally recruited the periaqueductal gray (PAG), a midbrain structure classically linked to threat processing and aversive learning. PAG activity closely followed model-derived signatures of reversal learning, implicating it in the suppression of previously rewarded actions and in updating behavior in the absence of explicit punishment. These findings reveal a previously unrecognized computational role for the PAG in value-based decision-making and cognitive flexibility, and substantiate task-fMRI as a powerful tool to study the rodent brain at a mesoscale resolution.

Introduction

Decision-making often occurs under conditions of uncertainty, yet individuals are generally capable of making decisions that lead to beneficial outcomes. This capacity relies on neural mechanisms that support the estimation of outcomes and the adjustment of behavior in response to changing environmental conditions, a process referred to as cognitive flexibility (Diamond, 2013). Reversal learning paradigms are prominently used to study cognitive flexibility across species (Highgate & Schenk, 2020; Izquierdo et al., 2017), in which switching of contingencies creates a violation of learned expectancies, resulting in rapid changes of behavioral responses. This requires the ability to detect unexpected outcomes, suppress previously reinforced responses, and update value representations in the light of new contingencies. These adaptive processes are supported by reinforcement learning mechanisms in the brain.

Reinforcement learning refers to the process by which organisms learn to associate specific stimuli or actions with rewarding or punishing outcomes, adapting their behavior to maximize positive results and minimize negative ones (Dayan & Niv, 2008; O'Doherty et al., 2017). A fundamental pathway supporting reinforcement learning is the dopaminergic input from the ventral tegmental area (VTA) to the ventral striatum, and specifically the nucleus accumbens, which serves as a teaching signal and plays a critical role in motivation and value-based decision-making (Barto, 1995; Glimcher, 2011; Montague et al., 1996; O'Doherty, 2004; Salamone & Correa, 2012; Schultz et al., 1997). Yet this mesolimbic circuitry does not operate in isolation. Converging evidence shows that other brain regions, including the prefrontal cortex, hippocampus, amygdala, and potentially even the periaqueductal gray (PAG), shape how reinforcement-learning related signals are generated, interpreted, and used (Ballard et al., 2019; Lee et al., 2012; Muller et al., 2024; Paton et al., 2006; Roy et al., 2014). Fully characterizing the circuitry involved in these behaviors therefore requires whole-brain approaches that can capture how canonical reward pathways interact with systems traditionally studied in other contexts (e.g., in aversion and defense).

Considerable understanding of brain-wide circuits contributing to reinforcement learning processes has come from human fMRI studies, where computational modeling of trial-by-trial neural signals has proven invaluable in dissecting the mechanisms by which brain regions encode learning, predict behavior, and interact within functional networks (Daw et al., 2006; Niv, 2009;

O'Doherty et al., 2003, 2007). Yet, despite their insight, human studies are inherently limited in their ability to establish causality, perform precise circuit manipulations, or achieve the level of experimental control possible in rodents. Extending this whole-brain imaging approach to behaving rodents therefore offers a unique opportunity to link systems-level network dynamics with cellular and molecular mechanisms of learning, bridging a critical translational gap. Until recently, most rodent imaging studies focused on resting-state fMRI to map functional connectivity across the mouse brain, further linking it with behavior (Bergmann et al., 2020; Grandjean et al., 2017; Lichtman et al., 2021; Liska et al., 2018). Recent advances by us and others, including the development of awake imaging setups and rodent-specific hemodynamic models (Bergmann et al., 2025; Desai et al., 2011; Fonseca et al., 2020; Han et al., 2019; Kahn et al., 2011; Lawen et al., 2025; Winkelmeier et al., 2022), now enable the implementation of reliable task-based fMRI. These developments open the door to probing not only canonical reward regions but also underappreciated contributors, offering a systems-level perspective that is critical for understanding the distributed computations underlying reinforcement learning and behavioral flexibility, which can then be further investigated and manipulated using invasive tools available in rodents.

Here, we used task-based fMRI of behaving mice to characterize the neural processes involved in a value-based decision-making task. We developed a non-invasive MR-compatible platform enabling high-resolution behavioral monitoring of head-fixed mice, which facilitated a longitudinal study of mice engaged in a go/no-go odor discrimination task followed by rule reversal, allowing us to investigate the distinct neural mechanisms engaged in initial acquisition versus reversal learning. We analyzed fMRI data using subject-specific trial-by-trial parameters derived from a reinforcement-learning model to better capture the temporal dynamics of learning, thereby improving predictive power and accounting for the small sample sizes feasible in rodent studies. This approach revealed the involvement of different brain regions in acquisition and reversal learning, including areas well-established as involved in goal-directed behavior like the nucleus accumbens, the dorsomedial striatum and the orbitofrontal cortex, but also, surprisingly, the PAG, which we found to be involved specifically in the reversal phase. Further examination showed that the PAG exhibits differential activity depending on correct behavioral outcome, is active during inhibition of lick responses to a "no-go" odor and is inactive during lick responses to a "go" odor.

Materials and Methods

Ethics

All animal experiments were conducted in accordance with the United States Public Health Service's Policy on Humane Care and Use of Laboratory Animals and approved by the Institutional Animal Care and Use Committee of the Technion—Israel Institute of Technology.

Animals and housing conditions

Twelve C57Bl/6J male mice (2–3 months old) were housed in groups of 2–5 animals per cage in a reversed 12-h light–dark cycle with food and water available *ad libitum* prior to water restriction. The housing room was maintained at 23 ± 2 °C. All experiments were conducted during the dark phase.

Head-post surgery

To minimize head movement during scanning, mice were implanted with MRI-compatible head posts, as previously described (Bergmann et al., 2025; Lichtman et al., 2021). Briefly, mice were anesthetized with isoflurane (1.5–2.5%), the scalp and periosteum were removed from above the surface of the skull, and a head post was attached to the skull using dental cement (C&B Metabond, Parkell, Brentwood, NY, United States). Mice received a subcutaneous injection containing broad-spectrum antibiotics (Cefalexin) and analgesia (Buprenorphine) during the surgery and daily for at least 3 days after the surgery, and were maintained in their home cage for a postoperative recovery period of 1 week.

MR-compatible behavioral setup

Experiments were conducted using an MRI-compatible behavioral platform designed to enable precise odor delivery, water reward control, and simultaneous recording of sniffing and licking behavior. The system included a custom head-fixation cradle, an air-dilution olfactometer for rapid odor presentation, the design of which has been previously described in detail (Arneodo et al., 2018; Shusterman et al., 2011), a non-invasive sniff sensor, a pressure-based lick detector, and a calibrated water-delivery mechanism. To adapt traditional olfactory setups for the MRI environment, the olfactometer was positioned outside the scanner bore with its output routed through a final solenoid valve located near the head-fixation apparatus. This configuration minimized magnetic interference while maintaining fast odor onset kinetics (steady-state concentration reached within ~100 ms). Sniffing was measured through a modified odor port connected to a miniature pressure transducer via short capillary tubing, providing stable, artifact-free respiration signals. Lick detection was achieved using a pressure-based MRI-compatible transducer, and water rewards (~2.5 µl per drop) were delivered through a solenoid-gated line controlled by an Arduino microcontroller. The olfactometer delivered a continuous flow of clean air (992 ml/min) that switched seamlessly to an odorized stream during stimulus presentation, preventing mechanical cues. Odorized air was generated by diverting nitrogen through odorant vials approximately 1 s before valve opening and mixing it into the main airflow to achieve a tenfold dilution. The final odor pulse lasted 1 s, after which clean air resumed. All flow paths were constructed from Teflon to prevent odor contamination. Odorant concentration profiles were verified using a photoionization detector (PID, Aurora Scientific, model 200B). All behavioral and control signals including sniffing, licks, valve triggers, and reward timing were synchronized and recorded via MATLAB-based scripts. This high-resolution monitoring approach provided precise temporal alignment between behavioral events and fMRI acquisition, ensuring accurate characterization of sensory, motor, and reward-related processes. For a more detailed description of the experimental setup, see Bergmann et al. (2025).

Flexible discrimination learning

Mice underwent an instrumental go/no-go odor discrimination task followed by a rule reversal in which odor contingencies were switched.

Following postoperative recovery, mice were single-housed and placed on a 10-day water restriction schedule, during which the amount of water was gradually reduced from 5 ml to 1 ml (1 ml per day). Water restriction was maintained throughout the experiment, including weekends. During pretraining (3–7 sessions), mice were first habituated to the setup to ensure acclimation to the scanner environment and were trained during mock scanning to lick a spout to obtain water rewards, with increasing inter-reward intervals (3–7 s), to shape stable licking behavior. Once mice consistently licked to consume water in the MRI setup and exhibited stable sniffing signals, they were scanned while performing the go/no-go task. All animals learned the task (i.e., there was no attrition).

The task consisted of multiple six-minute blocks, with 50 trials (25 per odor) per block, in which two neutral odorants (Pinene and Ethyl-Acetate) were presented pseudo-randomly. Each odor signaled either a rewarded (“go”) or unrewarded (“no-go”) condition. On each trial (2.5 s), an odor was presented for 1 s, and the animal was given a 2 s response window from odor onset. Inter-trial interval ranged from 5–12.5 s. Given a correct lick response to the go odor, the reward was delivered immediately. Lick responses to the no-go odor were not explicitly punished (only implicitly as water was not provided for these incorrect responses), and no lick responses were not rewarded.

During *Acquisition*, mice (n = 12) underwent 8–16 sessions (**Supplementary Table 1**) in which they learned to produce a lick response to the go odor and to withhold licking to the no-go odor. The first *Acquisition* session included 10–15 min of lick training during scanner calibration, allowing acquisition of fMRI data of the first odor presentation. In subsequent sessions, 1–2 blocks of the task were performed before fMRI data acquisition to allow for scanner calibration. These blocks were included in the behavioral data presented in **Figure 1** but were not included in the fMRI analyses in subsequent figures.

Next, a subset of this cohort (n = 6) underwent a *Reversal* phase for 5–6 sessions (**Supplementary Table 1**), during which they learned to lick in response to the presentation of the previously acquired no–go odor and to suppress the previously acquired lick response to the go odor. The *Reversal* group was randomly selected and showed learning performance similar to that of the remaining *Acquisition*-only mice (see results section). In the first *Reversal* session, mice performed 1–2 blocks of the original odor discrimination test during scanner calibration to ensure fMRI data acquisition for the first *Reversal* block. Subsequent sessions were performed using only the *Reversal* task, for which the first 1–2 blocks started before the fMRI data were acquired and therefore contributed to the behavioral data presented in **Figure 1**, but not to fMRI analyses in subsequent figures. Odor identity (go vs. no-go in the *Acquisition* phase) was counterbalanced across animals.

Reinforcement learning model

In order to model reinforcement learning we used a previously described (Nicholas et al., 2024) variant of a model-free Q-learning algorithm (Rescorla, 1972; Sutton & Barto, 1998). The model assumes a stored value, $Q(odor, action)$, for choosing an action of licking or not licking in response to a given odor. After each outcome, r_t, the Q value for the chosen action was updated according to:

$$Q(odor, action)_{t+1} = Q(odor, action)_t + \alpha(r_t - Q(odor, action)_t)$$

where the degree of updating was controlled by the learning rate α (**Supplementary Fig.1A**), which was a free parameter that ranges between 0 and 1. *Q* values of unchosen actions and other odors remained unchanged.

The model learned separate *Q* values for each odor (go, no-go, and control (nitrogen flow without odorant)) and action combination, such that six *Q* values were estimated in total, three for licking (*lick*) and three for not licking (*nolick*). The two *Q* values corresponding to the present stimulus, *o*, were used to compute a decision variable of the subject's response on each trial:

$$\Delta V_t = Q(o, lick)_t - Q(o, nolick)_t$$

The probability of licking was modeled using a logistic function:

$$P(lick) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \Delta V_t)}}$$

where β_0 is an intercept parameter that accounts for the bias towards licking and β_1 is an inverse temperature parameter that estimates the sensitivity to the learned values of actions related to the presented odor (**Supplementary Fig.1B**).

Model Fitting

We estimated model parameters for each animal using hierarchical Bayesian inference in order to allow group-level priors to regularize subject-level estimates. This approach to fitting reinforcement learning models improves parameter identifiability and predictive accuracy (van Geen & Gerraty, 2021) and has been used to fit similar Q-learning models (Nicholas et al., 2022). We first split the subjects into two groups: one group that underwent only acquisition (*n* = 6) and another that underwent a rule reversal following acquisition (*n* = 6). This splitting was performed because we reasoned that pooling data together from all animals would artificially inflate the learning rate of animals in the acquisition-only group. This is because animals in the reversal group experienced more trials where new learning was required, effectively doubling the "volatility" of this environment (Behrens et al., 2007).

To fit the model, the joint posterior was approximated using No-U-Turn Sampling (Hoffman & Gelman, 2014) as implemented in Stan (Carpenter et al., 2017). Four chains with 2000 samples

(1000 discarded as burn-in) were run for a total of 4000 posterior samples per model. Chain convergence was determined by ensuring that the Gelman-Rubin statistic, $\hat{R}$, was close to 1.

The model's likelihood function can be written as:

$$c_{s,t} \sim Bernoulli(\theta_{s,t})$$

where $c_{s,t}$ is 1 if subject s chose to lick on trial t and 0 if the subject chose not to lick, and $\theta_{s,t}$ is the estimated probability of this subject licking on this same trial. Following the recommendations of (Gelman & Hill, 2006), each subject's intercept and inverse temperature β_s were drawn from a multivariate normal distribution with mean vector μ_β and covariance matrix Σ_β:

$$\beta_s \sim MultivariateNormal(\mu_\beta, \Sigma_\beta)$$

where Σ_β was decomposed into a vector of coefficient scales τ_β and a correlation matrix Ω_β via:

$$\Sigma_\beta = diag(\tau_\beta) \times \Omega_\beta \times diag(\tau_\beta)$$

We set weakly-informative hyperpriors on the group-level hyperparameters μ_β, Ω_β and τ_β:

$$\mu_\beta \sim \mathcal{N}(0,5)$$

$$\tau_\beta \sim Cauchy^+(0,2.5)$$

$$\Omega_\beta \sim LKJCorr(2)$$

Each subject's learning rate parameters were also fit hierarchically with the following prior and hyperpriors (*a1, a2*):

$$\begin{gathered} \alpha_s \sim Beta(a1, a2) \\ a1 \sim \mathcal{N}(0,5) \\ a2 \sim \mathcal{N}(0,5) \end{gathered}$$

A description for why these prior and hyperpriors were chosen, as well as further details about the parameterization, can be found in Nicholas et al. (2022).

Image acquisition and preprocessing

MRI data were acquired as previously described (Bergmann et al., 2025). In brief, scans were acquired with a 9.4 Tesla MRI (Bruker BioSpin, Ettlingen, Germany) using a quadrature 86 mm transmit-only coil (Bruker BioSpin) and a 20 mm loop receive-only coil (Bruker BioSpin), and were reconstructed using ParaVision 5.1 (Bruker). Mice underwent multiple sessions of event-related fMRI (6–14 sessions during *Acquisition*, 5–6 sessions during *Reversal*), with each session containing multiple runs (2–12 runs per session), i.e., task blocks (see Flexible discrimination learning section). The rapid-presentation event-related design was generated using Optseq2 (Greve, 2002).

Mice were anesthetized for a short period of time (5% isoflurane) at the start of each session to allow for positioning them in the scanner. For each session, a relaxation enhancement (RARE) T2-weighted structural imaging (50 coronal slices, TR/TE 2300/8.5 ms, RARE factor = 4, flip angle = 180°, 200 × 200 × 300 μm^3, field of view of 19.2 × 19.2 mm^2, matrix size of 96 × 96) was first acquired while the mice performed the odor discrimination task. Then, blood oxygenation-level dependent (BOLD) contrast run scans were acquired for six minutes using spin echo-echo planar imaging (SE-EPI) sequence (TR/TE 2500/13.022 ms, flip angle = 90°, 50 coronal slices, voxel size 200 ×200× 300 μm^3, field of view of 14.4 × 9.6 mm^2, matrix size of 72 × 48). Preprocessing of raw data included removal of the first two volumes for T1-equilibration effects, compensation for slice-dependent time shifts, rigid body correction for head motion, semi-automatic linear registration (FSL FLIRT) to the Allen Mouse Brain Common Coordinate Framework version 3 (CCFv3, Kuan et al., 2015; Lein et al., 2007) that included a manual correction step for each session to validate proper alignment, and spatial smoothing with a full width at half maximum (FWHM) of 500 µm.

fMRI data analysis

fMRI data were analyzed using SPM12 (Wellcome Department of Cognitive Neurology, London, UK) and SnPM13 (http://nisox.org/Software/SnPM13/). The design matrices of all general linear models (GLM) computed in this study included the following nuisance regressors: global signal, ventricles signal, six motion parameters and their first-order derivatives, run constant for all runs excluding the last run, and events with frame displacement larger than the voxel size (200 µm).

Whole-brain analysis

In order to detect brain regions involved in tracking the value of an action (lick/no-lick) in a given state, we performed a whole-brain analysis using the decision variable (see Reinforcement-learning model section) as a covariate. For each animal, all sessions of the relevant experimental stage (Acquisition/Reversal) were combined to generate a single GLM in which events were defined as stimulus onset with a duration of zero. In addition, a decision variable regressor was created as a covariant by convolving the decision variable values with the mouse hemodynamic response function (HRF) previously modeled by our group (Bergmann et al., 2025). Due to the relatively slow TR (2.5 s) and the short dynamics of the mouse HRF, it was not possible to disentangle the stimulus, choice and feedback, and therefore, events were defined as single whole trials. A linear contrast of regressor coefficients was computed at the single-subject level for the decision variable regressor and was further used for a second, group-level analysis using a one-sample *t*-test. All group-level statistical maps were corrected for multiple comparisons using family-wise error correction. For the group-level parametric analysis ($n = 12$; **Figure 2**), the voxel extension was set to 5 voxels. For non-parametric statistical maps ($n = 6$; **Figure 3**), a permutation test was performed using the sign-flip approach in which 64 permutations, the equivalent of $2^{n_{\text{subject}}}$, were computed. To allow for cluster-level inference, we defined the variance smoothing to be the same FWHM that was applied to the data as instructed by SnPM13 manual. The cluster-defining threshold was set to t statistic of 6 ($\sim p < 0.001$, df = 5, prior to cluster-level inference), resulting in a critical STCS (suprathreshold cluster size) of 3 voxels.

Brain regions identified in the whole-brain analysis but located near fiber tracts, MRI artifacts, or close to the brain's boundaries were excluded from the results section.

Region of interest (ROI) analysis

Following the whole-brain findings of brain regions that are correlated to the Q-learning computation, we wanted to further assess the contribution of each of these regions to specific cognitive processes. Thus, for each experimental stage (*Acquisition* and *Reversal*), we defined each trial as an event based on the subject's response (Hit, False Alarm [FA], Correct Rejection [CR], and Miss) and generated a GLM per subject for each session separately with regressors corresponding to the different event types. We created specific ROI masks for regions identified based on the overlap between contiguous clusters of voxels in the statistical parametric/non-parametric maps and the Allen Mouse Brain Connectivity (AMBC) atlas (**Supplementary Fig. 3**). We used the MarsBaR toolbox (Brett et al., 2002) to extract finite impulse responses, plotting the hemodynamic response without assumptions on its response characteristics, with the onsets shifted by 5 s to allow to observe the pre-trial baseline and the full evolution of the response. We note that while the GLM used to identify the regions and the ROIs used in the ROI analysis are not independent, the inferences derived from this analysis are intended to characterize how Hits, FAs, CRs and Misses contribute to the response first identified in the Q-learning GLM which modeled the go and no-go conditions with the estimated decision variable as a parametric regressor. The ROI analyses therefore serve to quantify the fMRI response at both the individual-animal level and across these conditions.

Statistical analyses

Behavioral data were analyzed using MATLAB R2023 (The Mathworks, Natick, MA, USA), except for analysis of variance (ANOVA) which was performed using jamovi 2.3. For all repeated measures ANOVA tests, time was defined as a within-subject factor and the parametric tests were corrected for sphericity violation using the Huynh-Feldt method. For fMRI data, repeated measures ANOVA was performed in R (R Core Team, 2024) using RStudio (RStudio Team, 2020), with sphericity violation corrected using the Greenhouse-Geisser method.

Results

We utilized an experimental setup that allows for the delivery of odors into the scanner alongside a closed-loop system for lick detection and water delivery, allowing us to perform behavioral experiments with high precision in head-fixed mice (for a full description of the setup, see Bergmann et al., 2025). In this study, mice (n = 12) learned to perform an instrumental odor discrimination task while undergoing fMRI scanning (**Figure 1A**), with a subset of mice (n = 6) undergoing a rule reversal phase during which the action-outcome contingencies were switched.

Mice learn action-outcome associations and subsequent rule reversal in a go/no-go odor discrimination task

During task *Acquisition*, mice learned to discriminate between the rewarded stimulus (S+) and the unrewarded stimulus (S-) and reached a performance criterion of >80% correct responses by the fourth day (**Figure 1B**). A repeated-measures analysis of variance (rm-ANOVA) revealed a significant effect of Time ($F(5,11) = 25.2$, $p < 0.001$, $\eta^2 = 0.585$) as mice gradually learned the action-outcome associations for both stimuli. Evaluation of the reaction time from odor delivery to a lick response showed a decrease throughout sessions as animals became more proficient in the task (**Figure 1C**). Next, after the mice reached the criterion and maintained it, task contingencies were switched in the *Reversal* stage. Mice rapidly learned the new contingencies (**Figure 1D**), reaching a performance criterion on average within two sessions (Friedman rm-ANOVA; χ^2 (4,6) = 16.8, $p < 0.01$). The ability to quickly adapt to the switch in contingencies can be further observed by a decrease in reaction time throughout sessions during the *Reversal* phase (**Figure 1E**). Overall, mice presented goal-directed behavior during the experiment as indicated by the gradual lick responses to rewarded stimuli only (without using punishment for incorrect responses), as can be observed by a significant difference in reaction time for the first session of reversal relative to acquisition (**Figure 1F**; Wilcoxon Signed-rank test; W = 20, $p < 0.05$). Importantly, when evaluating for differences between the sub-group of mice that underwent only the acquisition phase and that who underwent a following reversal phase, we found no significant differences in learning the initial action-outcome association (two-way rm-ANOVA, $F(1,11) = 0.231$, $p = 0.947$, $\eta^2 = 0.006$).

Next, we used a reinforcement-learning model in order to parameterize the value of an action to a given odor. Specifically, we used a model-free Q-learning algorithm to allow for trial-by-trial estimation of action values (Q) of each odor to assess the process of choosing an action based on prior experiences. As estimated by the model, the probability of choosing to lick for the go odor accurately captures the experimental data (**Figure 1G**), showing a sharp decrease in probability at the point of reversal with a steep recovery. To illustrate Q-learning estimation, we plotted the decision variable values for one of the mice that participated in both *Acquisition* and *Reversal* (**Figure 1H**), demonstrating that the majority of learning occurred within approximately 500 trials in both stages (the equivalent of approximately two sessions), matching the time required to reach criterion for this animal. Similar estimation was observed for the other mice (not shown).

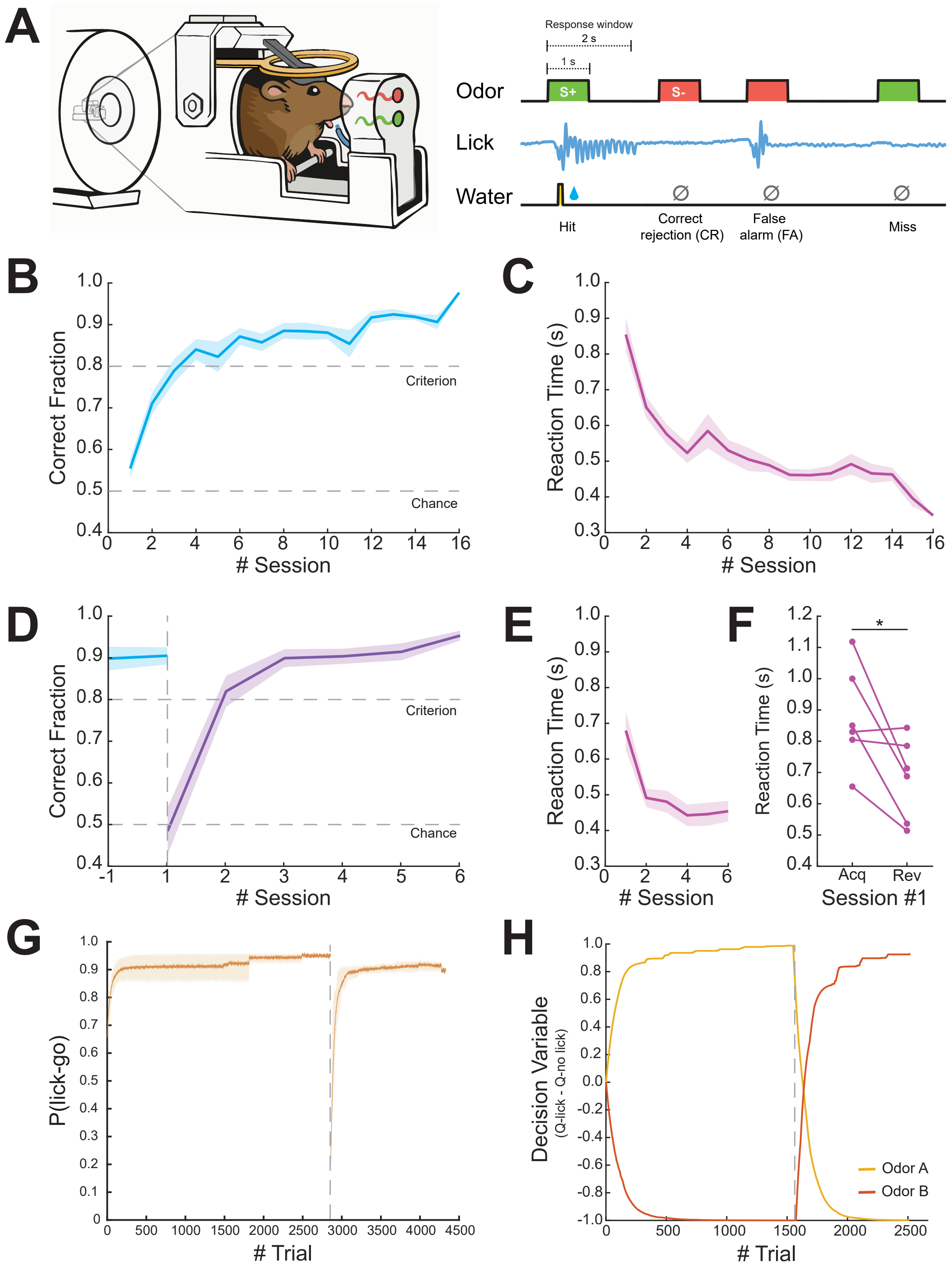

A
Response window
2 s
1 s
Odor
S+
S-
Lick
Water
Hit
Correct rejection (CR)
False alarm (FA)
Miss
B
Correct Fraction
Criterion
Chance
Session
C
Reaction Time (s)
Session
D
Correct Fraction
Criterion
Chance
Session
E
Reaction Time (s)
Session
F
*
Reaction Time (s)
Acq
Rev
Session #1
G
P(lick-go)
# Trial
H
Decision Variable
(Q-lick - Q-no lick)
Odor A
Odor B
Trial

Figure 1. Mice display flexible discrimination learning in a Go/No-go task in an experimental setup allowing for task-based fMRI. (A) Illustration of behavioral setup and task design. Odors were presented pseudo-randomly for 1 second with a response window lasting 2 seconds from odor onset. **(B)** Learning curve showing the averaged ratio of correct responses (Hit and Correct Rejection events) during task acquisition (*n* = 12). **(C)** Averaged reaction time of lick responses to Hit trials as a function of session. **(D)** Learning curve showing averaged ratio of correct responses (*n* = 6). Blue line indicates the last acquisition session prior to reversal. Purple line indicates the reversal phase. **(E)** Averaged reaction time of lick responses to Hit trials as a function of sessions during the reversal phase. **(F)** Pairwise comparison of the averaged reaction time for the first session in acquisition vs. reversal. **(G)** Averaged probability of choosing to lick in response to 'go' trials (*n* = 6). The gray dashed line indicates the rule reversal onset (95% $CI_{\text{acquisition}}$ = [0.580, 0.966], 95% CI_{reversal} = [0.157, 0.942]). **(H)** Decision variable ($Q_{\text{lick}} - Q_{\text{no-lick}}$) parameter of a representative animal as computed by the Q-learning model throughout the experiment. The yellow line represents odor A and the orange line represents odor B. The gray dashed line indicates the time of rule reversal. Data are shown as mean ± SEM (B, C, D and E). $^*p < 0.05$.

Brain responses indicative of value-based decision-making in the mouse brain

We sought to characterize the neural responses that are involved in reinforcement learning during flexible discrimination learning. We used the Q-learning algorithm in order to detect brain regions that take part in computing the value of choosing an action (lick/no-lick) to a given odor stimulus. Regions that track the decision variable estimated by the model, therefore, reflect the dynamic nature of learning at the individual-animal level. Specifically, in this task, regions identified using the decision variable are implicated in learning odor-action associations: one odor signals that licking will result in water delivery, while the other odor signals that licking will not yield reward. Though incorrect responses are not explicitly punished, animals learn to avoid unrewarded actions.

To evaluate responses in the task acquisition stage, we used high-field fMRI (9.4T) to measure distributed brain activity from the naïve state to task proficiency. We entered the decision variable of each animal (*n* = 12) as a parametric modulator in first-level GLM analysis, then conducted a second-level group analysis, calculating a statistical parametric map of regions that correlate with the decision variable (**Figure 2**). We observed responses in the basal ganglia (nucleus accumbens [ACB]), dorsomedial striatum [DMS], dorsolateral striatum [DLS] and globus pallidus externus [GPe]), regions in the insular cortex (agranular insular area [AI]) related to rewards and learning the cue-reward association, regions related to odor processing (tenia tecta [TTd] and anterior olfactory nuclei [AON]), regions related to avoidance and stress regulation (bed nucleus of the stria terminalis [BST]), and regions related water consumption (posterior AI/gustatory cortex [GU]). The most prominent responses (numerically) were observed in the striatum (ACB and

DMS). Overall, this measure reflects responses in brain regions previously implicated in value-based decision-making.

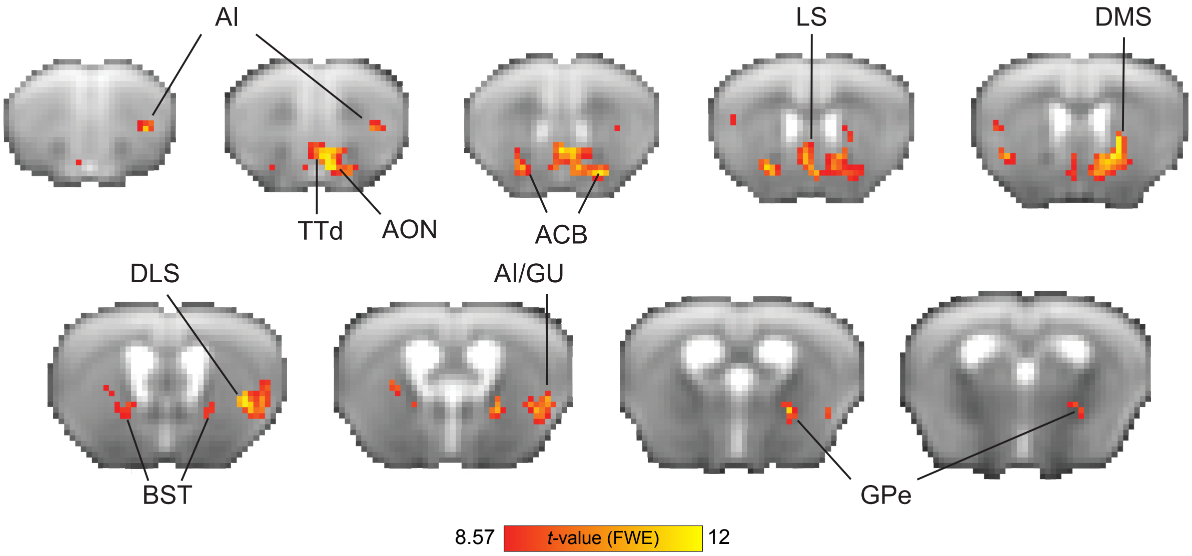


Figure 2. Neural substrates of Q-learning signals during go/no-go task acquisition. Group-level parametric one-sample *t* statistic maps showing BOLD response correlates of the decision variable values from the Q-learning model ($n = 12$ mice). The red color spectrum indicates areas with positive correlations to decision variable values, while blue colors indicate negative correlations. Maps are presented on the averaged raw fMRI data (spin-echo echo planar imaging) and annotated based on the Allen Mouse Brain Atlas; $p < 0.05$, corrected for multiple comparisons using family-wise error correction, voxel extension of 5. ACB, nucleus accumbens; AI, agranular insular area; AON, anterior olfactory nuclei; BST, bed nucleus of the stria terminalis; GU, gustatory cortex; DLS, dorsolateral striatum; DMS, dorsomedial striatum; GPe, globus pallidus externus; TT, tenia tecta; d, dorsal; v, ventral.

Next, we sought to characterize the neural responses during rule reversal. Reversal learning paradigms evaluate behavioral flexibility by switching the contingencies between stimuli and their outcomes, requiring subjects to change their learned responses when they encounter no reward for a previously rewarded response, as well as the ability to beneficially respond to a stimulus that was previously not reinforced. Using this experimental manipulation, we wanted to identify brain regions that are active during this cognitive process. Given that only a subset of mice underwent rule reversal, we used a non-parametric approach for whole-brain analysis and ran a permutation test. A GLM modeling the decision variable at the reversal stage of each animal ($n = 6$) was entered into a second-level group one-sample *t*-test, resulting in a statistical non-parametric map (**Figure 3**) of regions that were preferentially positively or negatively correlated with the value of licking or not licking, respectively. This map reveals the involvement of several regions that correlate with the decision variable values, showing a positive correlation in frontal and insular cortices (orbitofrontal [ORB] and AI/GU cortices), striatum (ACB and DMS), pallidum (medial septum [MS] and GPe) and olfactory processing regions (AON), and negative correlation in the dorsal hippocampus (HIPP) and ventral PAG.

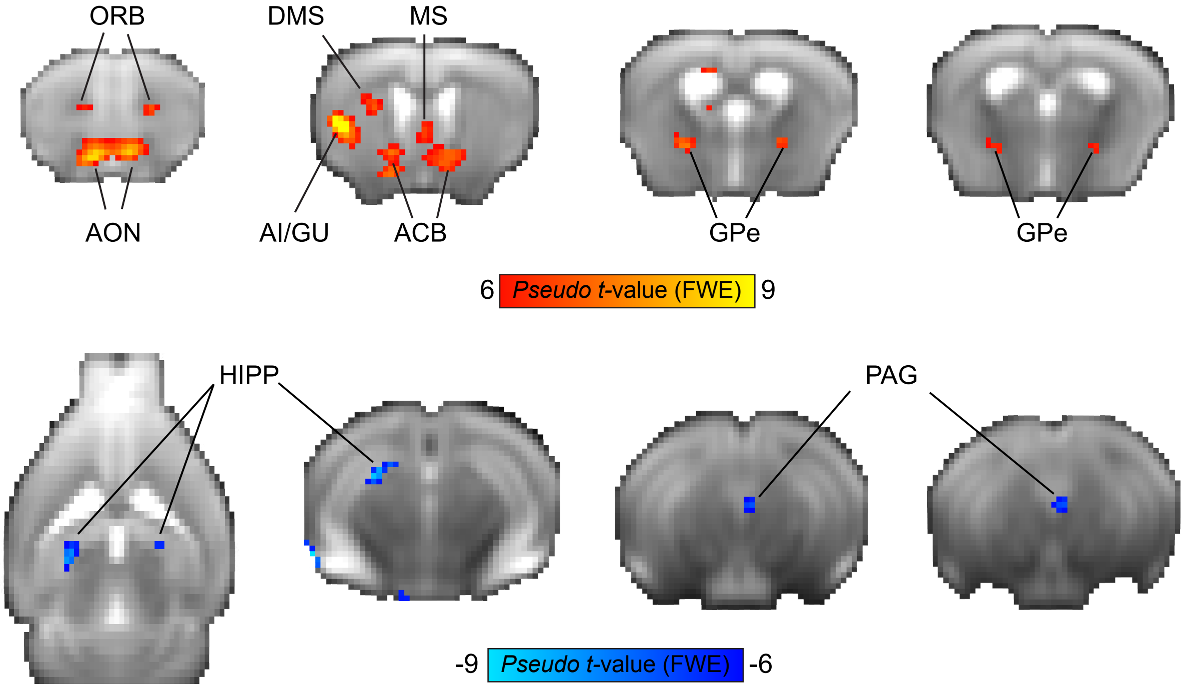


Figure 3. Neural substrates of Q-learning signals during reversal learning. Group-level non-parametric maps showing BOLD response correlates of the decision variable values from the Q-learning model ($n = 6$ mice). The red color spectrum indicates areas with positive correlations to the decision variable values, while blue colors indicate negative correlations. Maps are presented on an average raw fMRI data (spin-echo echo planar imaging) and annotated based on the Allen Mouse Brain Atlas; $p < 0.05$, corrected for multiple comparisons using family-wise error correction, voxel extension of 3. ACB, nucleus accumbens; AI, agranular insular area; AON, anterior olfactory nuclei; DMS, dorsomedial striatum; GPe, globus pallidus externus; GU, gustatory cortex; HIPP, hippocampus; MS, medial septum; ORB, orbitofrontal cortex; PAG, periaqueductal gray.

Activation of PAG correlates with beneficial behavioral responses in reversal learning only

Given recent findings suggesting that PAG represents an aversive prediction error, and the lack of evidence linking it to reversal learning and action value, especially under appetitive conditions, we next sought to characterize its responses in contrast to those in the ACB, a well-established region for prediction error computation. We therefore examined the PAG and ACB contributions to the different cognitive components of discrimination learning using an ROI analysis, and looked at finite impulse responses (FIR) for conditions differentiated by their behavioral responses, whether correct or incorrect, and their outcomes (Hit, Correct Rejection, False Alarm, and Miss; see **Supplementary Fig. 2** for behavioral performance).

We estimated FIR-based BOLD response time courses from the PAG and ACB and computed the mean area under the curve (AUC) across time points for each session to quantify overall response magnitude. Focusing on the PAG, a two-way repeated-measure ANOVA, with Condition (Hit, Correct Rejection, False Alarm, and Miss) and Session as within-subject factors, revealed a significant main effect of Condition ($F(1.77, 8.87) = 15.17$, $p < 0.01$, generalized $\eta^2 = 0.34$), indicating that AUC varied across behavioral outcomes, distinguishing correct from incorrect responses, and as can be seen in the analyses below, was driven mainly by positive responses to Correct Rejection and negative to Hit. In contrast, neither the main effect of Session nor the Session × Condition interaction was significant, suggesting that overall effects were stable across sessions, with learning occurring in the first session, consistent with the animals' behavioral outcomes. Analysis of AUC values from the ACB showed a significant main effect of Condition ($F(1.66, 8.30) = 20.61$, $p < .001$, generalized $\eta^2 = 0.55$) and a significant interaction of Session × Condition ($F(12, 60) = 2.20$, $p = .023$, generalized $\eta^2 = 0.19$), with no significant effect of Session.

Next, we compared responses between the PAG and ACB. Given that as the animal learns the decision variable stabilizes and is maximally different between correct licks for the go odor (Hit) and avoidance for the no-go odor (Correct Rejection) (**Figure 1H**), we focused on these two conditions during the last *Reversal* session when mice had already fully learned the rule reversal (**Figure 4**). A two-way repeated-measure ANOVA revealed a significant interaction of ROI × Condition ($F(1, 5) = 55.13$, $p < .001$, generalized $\eta^2 = 0.83$), indicating that the behavioral effect of Condition reversed between PAG and ACB. Post hoc pairwise comparison performed within each ROI demonstrated that ACB showed a strong activation to Hit events and inactivation to Correct Rejection (**Figure 4B,C**; bottom row) ($t(5) = 4.19$, $p = .020$, Holm-adjusted). Conversely, PAG showed inactivation to Hit events and activation to Correct Rejection (**Figure 4B,C**; top row; $t(5) = 8.95$, $p = 0.002$). Additional cross-region contrasts confirmed this double dissociation, showing that for Correct Rejection the AUC was higher in PAG than in ACB, whereas for Hit the AUC was higher in ACB than in PAG (both $p \leq 0.025$, Holm-adjusted). Taken together, these findings reveal a crossover interaction in which ACB and PAG exhibit opposite BOLD response modulations for Correct Rejection versus Hit conditions, highlighting distinct functional contributions of the two regions to outcome processing.

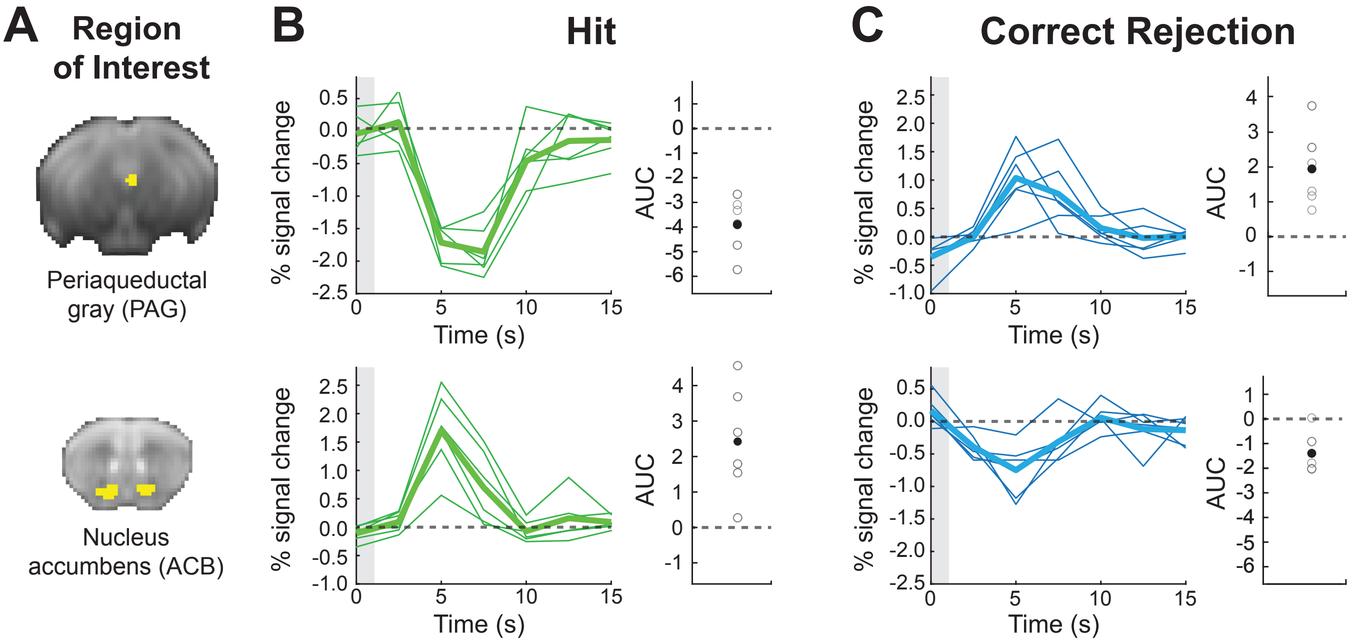


Figure 4. Opposing roles of the periaqueductal gray (PAG) and nucleus accumbens (ACB) in mediating adaptive behavior in proficient mice. Region of interest (ROI) analysis showing the time course fMRI BOLD response in PAG (*top*) and ACB (*bottom*) for the fifth session in the reversal stage. **(A)** ROI masks are presented on an average raw fMRI image (spin-echo echo planar imaging). **(B)** BOLD fMRI Responses to Hit (lick response to go trials; *green*) and **(C)** Correct Rejection (no lick to no-go trials; *blue*) are shown. The thick lines represent the group averaged response and the thin lines show individual animals. The gray boxes at time zero depict the odor stimulus timing (1 s). Group mean area under the curve (AUC) of the fMRI response (filled circles) and individual animals (open circles) demonstrate consistent responses at the group and individual animal levels.

Finally, we wanted to examine whether the activity observed in PAG is specific to reversal learning or whether this region contributes to acquisition learning as well (**Figure 5**). We extracted FIR responses for an *Acquisition* session at a timepoint corresponding to reversal learning, when mice demonstrated comparable performance levels (**Figure 5A**; Wilcoxon Signed-rank test; $p = 1$, $Z = 0$), and computed the AUC values for the two conditions. Looking at the Hit condition, we found a significant decrease in AUC values for *Acquisition* relative to *Reversal* (**Figure 5B**; Wilcoxon Signed-rank test; $p = 0.031$, $Z = 21$). Comparison of the Correct Rejection condition revealed a trend (**Figure 5C**; Wilcoxon Signed-rank test; $p = 0.093$, $Z = 2$), showing overall decrease in AUC values. Further, PAG responses during the *Acquisition* phase were not significantly different than baseline in both conditions (Sign test; Hit: $p = 0.218$, sign = 1; Correct Rejection: $p = 1$, sign = 3). Collectively, the results indicate that the subregion in PAG that was found to be important for learning of correct behavioral responses in *Reversal* does not seem to be important for *Acquisition*.

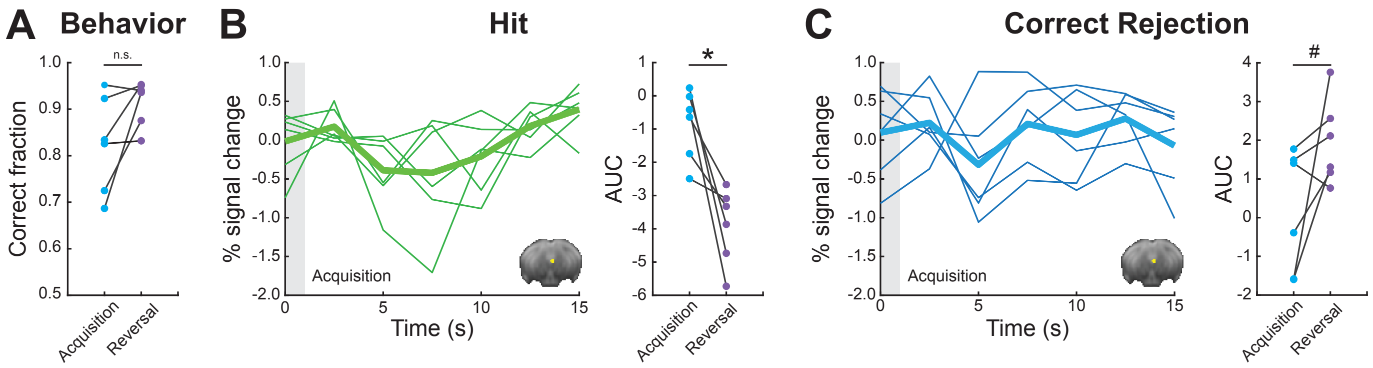


Figure 5. PAG responses do not contribute to adaptive behavior in proficient mice during Acquisition. (A) Pairwise comparison of behavioral performance during *Acquisition* vs. *Reversal* for sessions 4/5. **(B)** FIR responses for Hit condition during *Acquisition* (left). Pairwise comparison of AUC values showing individual animals for *Acquisition* vs. *Reversal* (right). **(C)** FIR responses for Correct Rejection condition during *Acquisition* (left). Pairwise comparison of AUC values showing individual animals for *Acquisition* vs. *Reversal* (right). For FIR responses, the thick lines represent group averages and the thin lines show individual animals. The gray boxes at time zero depict the odor stimulus timing (1 s). *$p < 0.05$, $^{\#}p < 0.1$, n.s. no significance. Insets show PAG region of interest mask presented on an average raw fMRI image (spin-echo echo planar imaging).

Discussion

Flexible, goal-directed behavior relies on the capacity to adapt action–outcome associations when contingencies change. This process, typically studied using reversal learning paradigms, is thought to depend on corticostriatal circuits and dopaminergic teaching signals, yet it remains unclear whether additional brain regions may also contribute. Whole-brain imaging approaches provide an opportunity to address this question. In this study, we found that the periaqueductal gray (PAG) contributes to cognitive flexibility by supporting the suppression of previously reinforced actions in the absence of explicit punishment, and assessed its contributions relative to the nucleus accumbens. By combining task-based fMRI with reinforcement learning modeling in mice performing an odor discrimination task, we found that acquisition was supported by the nucleus accumbens, whereas reversal learning additionally engaged the PAG. Notably, PAG responses contrasted with those in the nucleus accumbens, exhibiting preferential activation during correct rejection of no-go cues and suppression during correct approach to go cues. Together, these results identify the PAG as a key contributor to reversal learning, expanding current models of the neural mechanisms underlying behavioral flexibility. Moreover, they highlight the strength of whole-brain fMRI in rodents, demonstrating that novel findings can emerge even within a well-established and extensively studied behavioral paradigm.

By applying a model-free reinforcement-learning algorithm we were able to capture trial-by-trial dynamics of value updating and link them to fMRI responses. This computational approach was essential for detecting the emergence of PAG activity during reversal, as it allowed us to model how action values evolve across individual trials rather than relying on averaged performance measures. Computational modeling of reinforcement learning has been highly influential in human fMRI studies (Niv, 2009; O'Doherty et al., 2003), and its value has only recently been shown in rodent fMRI as well (Winkelmeier et al., 2022). Our results demonstrate the feasibility and utility of such models in animal neuroimaging, highlighting how model-based analyses can improve sensitivity to dynamic neural processes underlying learning. This methodological advance also paves the way for more direct cross-species comparisons of reinforcement learning circuitry.

Beyond the striatum, our whole-brain analyses revealed that acquisition engaged a distributed set of regions, including the agranular insula, gustatory cortex, anterior olfactory nuclei, and bed nucleus of the stria terminalis. These findings are consistent with prior reports that value-based learning recruits not only canonical reward regions but also areas involved in sensory processing, interoception, and stress regulation (FitzGerald et al., 2013; Ge & Balleine, 2022; Hernández-Ortiz et al., 2023; Kogan & Fontanini, 2024; Levinson et al., 2020). The engagement of olfactory and gustatory cortices likely reflects the multimodal nature of the task, while activity in the insula and bed nucleus of the stria terminalis may reflect the integration of reward signals with internal state and arousal. Together, these observations emphasize that flexible discrimination learning emerges from the interaction of distributed brain systems, with the PAG contributing selectively during reversal to bias action suppression once initial associations have been formed.

Classically, the PAG has been implicated in defensive behaviors (Bandler & Keay, 1996; Carrive, 1993), processing nociceptive signals (Basbaum & Fields, 1984; Behbehani, 1995), and coordinating autonomic responses to threat (Dampney, 1994; Keay & Bandler, 2001). However, there has been emerging evidence linking it to behavioral flexibility and value-based decision-making (Ozawa et al., 2017; Reis et al., 2021; Sukikara et al., 2006; Wright & McDannald, 2019). Recent work has highlighted PAG as a potential relay between brainstem value signals and forebrain decision circuits (Gorka et al., 2023; Roy et al., 2014). Further, PAG neurons were shown to encode both negative and positive prediction errors (Walker et al., 2020; Wright & McDannald, 2019), and project to thalamic and cortical areas implicated in strategy updating (Assareh et al., 2016; Faull et al., 2019; Kragel et al., 2019; Krout & Loewy, 2000). Thus, PAG serves as a hub that transforms aversive sensory input into adaptive motor and physiological outputs, thereby guiding rapid survival-related responses. While the majority of previous studies linking PAG to reinforcement learning processes used aversive, pain-related paradigms, our task design allowed for reversal learning to occur in the absence of explicit punishment. Mice adapted their behavior solely through the omission of expected reward, suggesting that PAG activity may contribute to updating value representations when contingencies change, even under neutral conditions. Our results extend the current framework by showing that PAG recruitment can occur in appetitive tasks without negative reinforcement, highlighting its broader role in signaling the need to suppress outdated responses. This observation aligns with a proposal that PAG contributes to the evaluation of approach versus avoidance strategies in changing environments (Tryon & Mizumori, 2018) and suggests it may act as a general mediator of adaptive response suppression.

The region of interest analysis revealed a striking double dissociation between PAG and nucleus accumbens activity during reversal learning. Whereas nucleus accumbens responses were strongest for rewarded licks (hits), PAG responses were selectively enhanced during correct rejections. This opposing response indicates that PAG and nucleus accumbens contribute complementary signals to support flexible decision-making. The nucleus accumbens has long been implicated in representing positive prediction errors and guiding approach behavior (Montague et al., 1996; Nicola, 2007; Schultz et al., 1997). By contrast, the PAG response profile suggests a role in reinforcing the suppression of actions that no longer yield reward, consistent with its established involvement in aversive learning (Johansen et al., 2010; Walker et al., 2020). This opponency suggests that midbrain and striatal circuits jointly encode both the drive to exploit rewarded contingencies and the need to avoid perseverating on outdated responses. Through its reciprocal connections with both brainstem neuromodulatory centers and midbrain dopamine neurons, the PAG is anatomically poised to influence the teaching signals that drive reinforcement learning even in conditions beyond those classically associated with it. Namely, avoidance learning could, at least in part, be mediated by PAG computations by biasing dopaminergic signaling toward actions in situations where no explicit punishment occurs.

Importantly, PAG responses were not observed during initial acquisition, even when behavioral performance was comparable to that achieved in reversal. This indicates that PAG recruitment is not a general feature of value-based learning, but rather emerges selectively when animals must overcome prior learning. The specificity of PAG involvement in reversal echoes prior rodent and primate studies implicating cortical circuits, the orbitofrontal cortex in particular, in behavioral flexibility (Cools et al., 2002; Ghahremani et al., 2010; Izquierdo et al., 2017; Schoenbaum et al., 2006). Our findings extend this literature by demonstrating that the PAG is also selectively engaged under reversal conditions. This supports the notion that flexible decision-making depends on coordinated contributions from both cortical and subcortical regions, with PAG providing a key midbrain computation to facilitate behavioral adaptation.

Collectively, our results expand the functional repertoire of the PAG beyond its established role in aversion and defensive behaviors, positioning it as a key node in the neural circuitry that supports flexible decision-making. By demonstrating that PAG activity is selectively recruited during reversal learning, and exhibits functional opponency with striatal reward signals, our findings suggest that PAG contributes to updating value representations when prior contingencies are no longer valid. This role is particularly notable given that mice adapted their behavior without explicit punishment, implying that PAG computations may bias action selection toward adaptive avoidance even in neutral contexts. More broadly, these results underscore the importance of brainstem–forebrain interactions in reinforcement learning and provide a systems-level framework for future studies examining how PAG signals integrate with dopaminergic and cortical circuits to support cognitive flexibility. Elucidating these mechanisms will be critical for understanding how distributed midbrain circuits contribute to adaptive behavior and how their dysfunction may contribute to neuropsychiatric disorders characterized by impaired behavioral flexibility.

Acknowledgments

This work was supported by the Israel Science Foundation (770/17; to I.K.), the National Institutes of Health (1R01NS091037; to I.K.), the Adelis Foundation (to I.K.) and the Prince Center (to I.K.). We thank Technion’s Biological Core Facilities and Edith Suss-Toby for her assistance with the MRI, the Technion Preclinical Research Authority and Nadav Cohen for assistance with animal care, Yael Niv for her helpful comments on the manuscript, and Matteo Farinella for his assistance in illustrating the behavioral setup shown in Figure 1A. We acknowledge the use of generative AI tools (ChatGPT, GPT-4.5, Open AI) to assist in manuscript editing, specifically for grammar and style improvements, as well as enhancing overall readability. These tools did not influence the scientific content or interpretation of our data. Conflict of Interest: None declared.

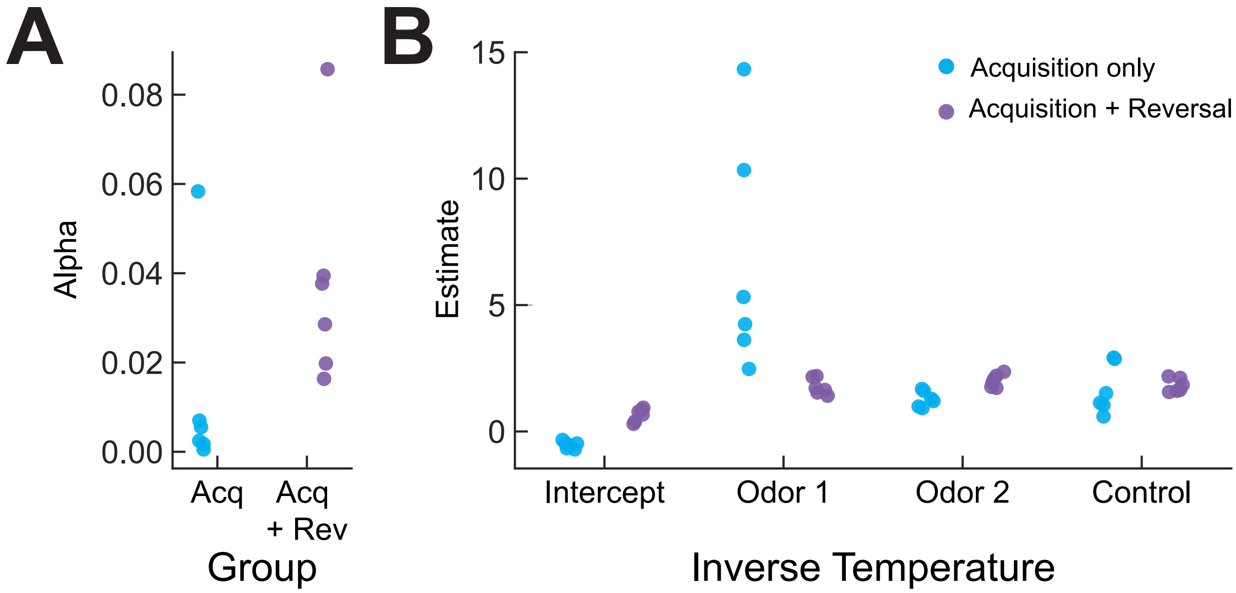
Author contributions

D.L. and I.K. designed the study. D.L. and E.B. conducted experiments. D.L., E.B. and J.N. wrote software for data analysis. R.T.G. and D.L. designed fMRI data analysis. D.L. analyzed the data. D.L. and I.K. prepared the manuscript.

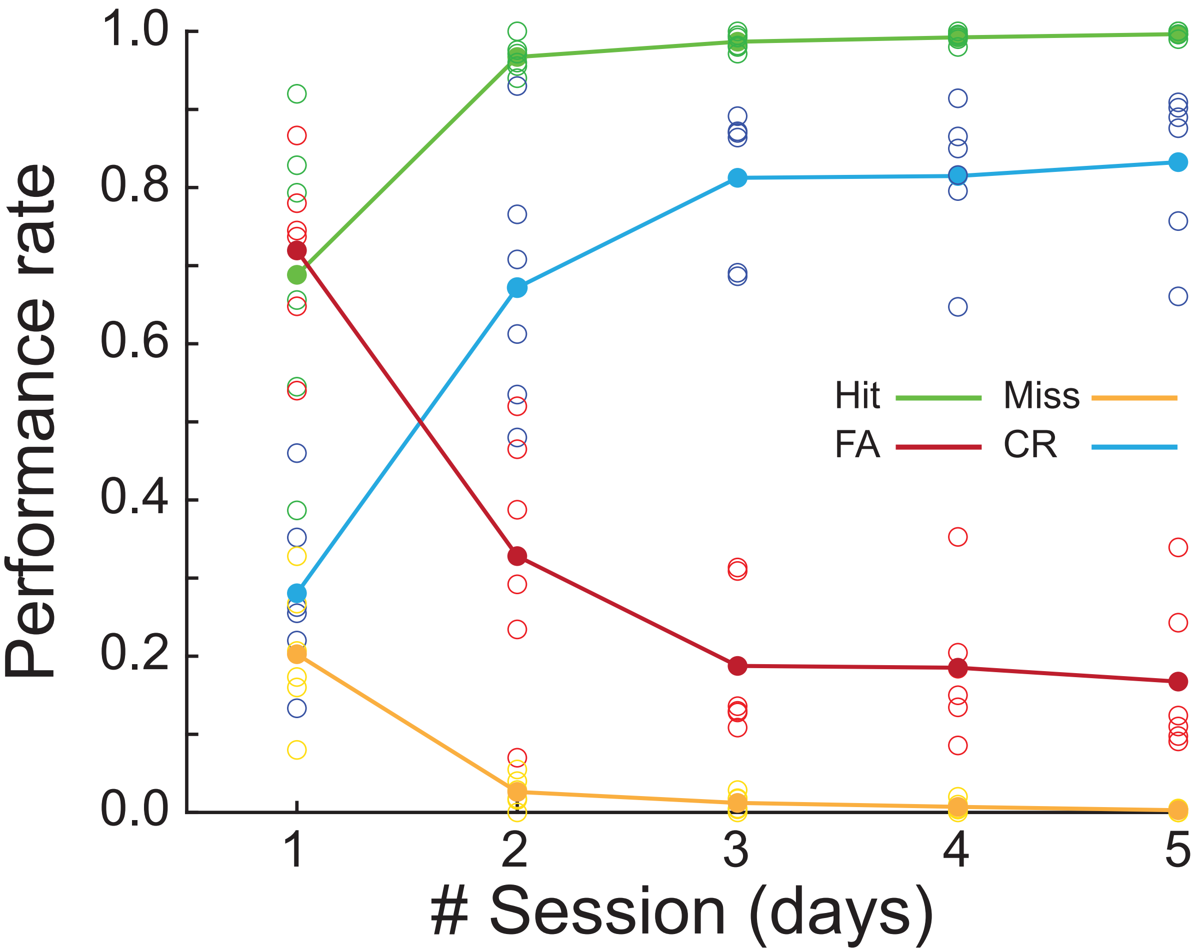
Data availability

Raw imaging data and all statistical maps included in this report will be made available in BIDS format on OpenNeuro upon publication.

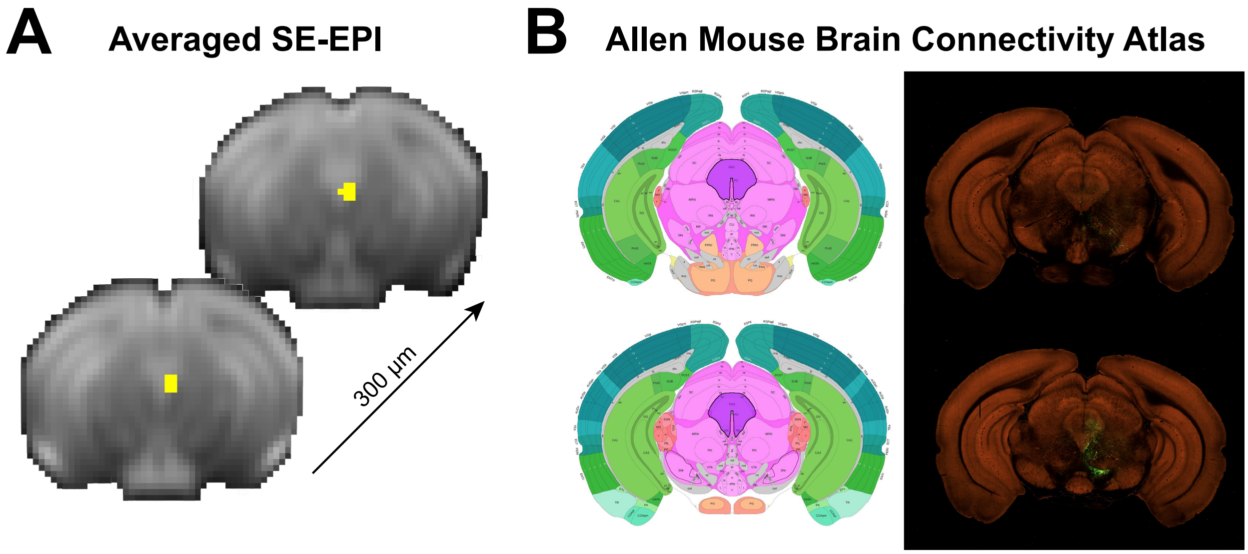
Supplementary Materials



Supplementary Figure 1. Group parameters computed by the Q-learning algorithm. Fitting parameters are shown for the two modeling groups: acquisition only (*blue*) and acquisition following reversal (*purple*; *n* = 6 per group). **(A)** Learning rate parameter alpha. **(B)** Beta estimates for intercept (bias to lick), odor 1 (go stimulus), odor 2 (no-go stimulus) and control for changes in airflow during final valve opening (nitrogen).



Supplementary Figure 2. Group performance for the different discrimination conditions during Reversal. Behavioral performance rates across five consecutive sessions during the Reversal phase divided by the lick response of the animal to each cue type: Hit (lick to go odor; *green*), Miss (no lick to go odor; *yellow*), False Alarm (FA, lick to no-go odor; *red*) and Correct Rejection (CR, no lick to no-go odor; *light blue*). Filled circles indicate group mean and open circles indicate individual subjects ($n = 6$).



Supplementary Figure 3. Spatial localization of the periaqueductal gray across fMRI and atlas space. **(A)** Periaqueductal gray mask (yellow) defined based on regions showing a significant BOLD response in the whole-brain analysis and further used in the ROI analysis. The ROI mask is presented on an average raw fMRI data (spin-echo echo planar imaging), shown as sequential coronal slices with a slice thickness of 300 μm. **(B)** Coronal images (*left*, atlas; *right*, two-photon tomography) taken from the Allen mouse brain connectivity atlas that correspond to the spatial location of the fMRI data. The region highlighted in purple denotes the Periaqueductal gray as defined by the Allen Institute. Image identification numbers are 87 (bottom) and 90 (top) in the reference atlas.

Supplementary Table 1. Summary of number of sessions completed by each subject.

Subject	**# Sessions**			
	Behavior		**fMRI**	
	Acquisition	*Reversal*	*Acquisition*	*Reversal*
Subject 01	10	—	9	—
Subject 02	10	—	9	—
Subject 03	6	—	6	—
Subject 04	9	5	8	5
Subject 05	15	—	15	—
Subject 06	16	6	14	6
Subject 07	15	6	13	6
Subject 08	14	6	12	6
Subject 09	9	—	9	—
Subject 10	9	—	9	—
Subject 11	8	5	8	5
Subject 12	8	5	8	5

Summary of maximal number of sessions (one session per day) completed by each subject during the *Acquisition* and *Reversal* experimental phases (Behavior). All subjects reached learning criterion in the *Acquisition* phase by the fourth session (3.583 ± 1.621, mean ± SD). Of the 6 subjects participating in the *Reversal* phase, 3 subjects completed 8–9 sessions, and 3 completed 14–16 sessions. This manipulation was used to rule out that an extended number of sessions during *Acquisition* affects the results observed during *Reversal*. The table also shows the number of usable fMRI data (fMRI) session, as some were excluded due to software issues or poor performance inside the scanner.

References

Arneodo, E. M., Penikis, K. B., Rabinowitz, N., Licata, A., Cichy, A., Zhang, J., Bozza, T., & Rinberg, D. (2018). Stimulus dependent diversity and stereotypy in the output of an olfactory functional unit. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-03837-1

Assareh, N., Sarrami, M., Carrive, P., & McNally, G. P. (2016). The organization of defensive behavior elicited by optogenetic excitation of rat lateral or ventrolateral periaqueductal gray. *Behavioral Neuroscience*, *130*(4). https://doi.org/10.1037/bne0000151

Ballard, I. C., Wagner, A. D., & McClure, S. M. (2019). Hippocampal pattern separation supports reinforcement learning. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-08998-1

Bandler, R., & Keay, K. A. (1996). Columnar organization in the midbrain periaqueductal gray and the integration of emotional expression. In *Progress in Brain Research* (Vol. 107). https://doi.org/10.1016/s0079-6123(08)61871-3

Barto, A. G. (1995). Adaptive critics and the basal ganglia. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia.* (pp. 215–232). The MIT Press.

Basbaum, A. I., & Fields, H. L. (1984). Endogenous pain control systems: Brainstem spinal pathways and endorphin circuitry. *Annual Review of Neuroscience*, *VOL. 7*. https://doi.org/10.1146/annurev.ne.07.030184.001521

Behbehani, M. M. (1995). Functional characteristics of the midbrain periaqueductal gray. In *Progress in Neurobiology* (Vol. 46, Issue 6). https://doi.org/10.1016/0301-0082(95)00009-K

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9). https://doi.org/10.1038/nn1954

Bergmann, E., Gofman, X., Kavushansky, A., & Kahn, I. (2020). Individual variability in functional connectivity architecture of the mouse brain. *Communications Biology*, 1–10. http://dx.doi.org/10.1038/s42003-020-01472-5

Bergmann, E., Lichtman, D., Resulaj, A., Yona, G., Nahman, O., Rinberg, D., & Kahn, I. (2025). Striatal and hippocampal contributions to instrumental learning in the mouse using high-resolution behavioral monitoring and fMRI. *Imaging Neuroscience*, *3*. https://doi.org/10.1162/IMAG.a.975

Brett, M., Anton, J. L., Valabregue, R., & Poline, J. B. (2002). Region of interest analysis using an SPM toolbox. *NeuroImage*, *16*.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1). https://doi.org/10.18637/jss.v076.i01

Carrive, P. (1993). The periaqueductal gray and defensive behavior: Functional representation and neuronal organization. *Behavioural Brain Research*, *58*(1–2). https://doi.org/10.1016/0166-4328(93)90088-8

Cools, R., Clark, L., Owen, A. M., & Robbins, T. W. (2002). Defining the Neural Mechanisms of Probabilistic Reversal Learning Using Event-Related Functional Magnetic Resonance Imaging. *Journal of Neuroscience*, *22*(11). https://doi.org/10.1523/jneurosci.22-11-04563.2002

Dampney, R. A. L. (1994). Functional organization of central pathways regulating the cardiovascular system. In *Physiological Reviews* (Vol. 74, Issue 2). https://doi.org/10.1152/physrev.1994.74.2.323

Daw, N. D., O’Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095). https://doi.org/10.1038/nature04766

Dayan, P., & Niv, Y. (2008). Reinforcement learning: The Good, The Bad and The Ugly. In *Current Opinion in Neurobiology* (Vol. 18, Issue 2). https://doi.org/10.1016/j.conb.2008.08.003

Desai, M., Kahn, I., Knoblich, U., Bernstein, J., Atallah, H., Yang, A., Kopell, N., Buckner, R. L., Graybiel, A. M., Moore, C. I., & Boyden, E. S. (2011). Mapping brain networks in awake mice using combined optical neural control and fMRI. *Journal of Neurophysiology*, *105*(3). https://doi.org/10.1152/jn.00828.2010

Diamond, A. (2013). Executive functions. In *Annual Review of Psychology* (Vol. 64). https://doi.org/10.1146/annurev-psych-113011-143750

Faull, O. K., Subramanian, H. H., Ezra, M., & Pattinson, K. T. S. (2019). The midbrain periaqueductal gray as an integrative and interoceptive neural structure for breathing. In *Neuroscience and Biobehavioral Reviews* (Vol. 98). https://doi.org/10.1016/j.neubiorev.2018.12.020

FitzGerald, T. H. B., Friston, K. J., & Dolan, R. J. (2013). Characterising reward outcome signals in sensory cortex. *NeuroImage*, *83*. https://doi.org/10.1016/j.neuroimage.2013.06.061

Fonseca, M. S., Bergomi, M. G., Mainen, Z. F., & Shemesh, N. (2020). Functional MRI of large scale activity in behaving mice. *BioRxiv*.

Ge, M., & Balleine, B. W. (2022). The role of the bed nucleus of the stria terminalis in the motivational control of instrumental action. *Frontiers in Behavioral Neuroscience*, *16*. https://doi.org/10.3389/fnbeh.2022.968593

Gelman, A., & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press. https://doi.org/10.1017/CBO9780511790942

Ghahremani, D. G., Monterosso, J., Jentsch, J. D., Bilder, R. M., & Poldrack, R. A. (2010). Neural components underlying behavioral flexibility in human reversal learning. *Cerebral Cortex*, *20*(8). https://doi.org/10.1093/cercor/bhp247

Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(SUPPL. 3). https://doi.org/10.1073/pnas.1014269108

Gorka, A. X., Philips, R. T., Torrisi, S., Manbeck, A., Goodwin, M., Ernst, M., & Grillon, C. (2023). Periaqueductal gray matter and medial prefrontal cortex reflect negative prediction errors during differential conditioning. *Social Cognitive and Affective Neuroscience*, *18*(1). https://doi.org/10.1093/scan/nsad025

Grandjean, J., Zerbi, V., Balsters, J. H., Wenderoth, N., & Rudin, M. (2017). Structural Basis of Large-Scale Functional Connectivity in the Mouse. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *37*(34), 8092–8101. https://doi.org/10.1523/JNEUROSCI.0438-17.2017

Greve, D. N. (2002). *optseq2*. Optseq Home Page.

Han, Z., Chen, W., Chen, X., Zhang, K., Tong, C., Zhang, X., Li, C. T., & Liang, Z. (2019). Awake and behaving mouse fMRI during Go/No-Go task. *NeuroImage*, *188*. https://doi.org/10.1016/j.neuroimage.2019.01.002

Hernández-Ortiz, E., Luis-Islas, J., Tecuapetla, F., Gutierrez, R., & Bermúdez-Rattoni, F. (2023). Top-down circuitry from the anterior insular cortex to VTA dopamine neurons modulates reward-related memory. *Cell Reports*, *42*(11). https://doi.org/10.1016/j.celrep.2023.113365

Highgate, Q., & Schenk, S. (2020). Cognitive flexibility in humans and other laboratory animals. In *Journal of the Royal Society of New Zealand*. https://doi.org/10.1080/03036758.2020.1784240

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. In *Journal of Machine Learning Research* (Vol. 15). http://mcmc-jags.sourceforge.net

Izquierdo, A., Brigman, J. L., Radke, A. K., Rudebeck, P. H., & Holmes, A. (2017). The neural basis of reversal learning: An updated perspective. In *Neuroscience* (Vol. 345, pp. 12–26). Elsevier Ltd. https://doi.org/10.1016/j.neuroscience.2016.03.021

Johansen, J. P., Tarpley, J. W., Ledoux, J. E., & Blair, H. T. (2010). Neural substrates for expectation-modulated fear learning in the amygdala and periaqueductal gray. *Nature Neuroscience*, *13*(8). https://doi.org/10.1038/nn.2594

Kahn, I., Desai, M., Knoblich, U., Bernstein, J., Henninger, M., Graybiel, A. M., Boyden, E. S., Buckner, R. L., & Moore, C. I. (2011). Characterization of the Functional MRI Response

Temporal Linearity via Optical Control of Neocortical Pyramidal Neurons. *Journal of Neuroscience*, *31*(42), 15086–15091. https://doi.org/10.1523/JNEUROSCI.0007-11.2011

Keay, K. A., & Bandler, R. (2001). Parallel circuits mediating distinct emotional coping reactions to different types of stress. In *Neuroscience and Biobehavioral Reviews* (Vol. 25, Issues 7–8). https://doi.org/10.1016/S0149-7634(01)00049-5

Kogan, J. F., & Fontanini, A. (2024). Learning enhances representations of taste-guided decisions in the mouse gustatory insular cortex. *Current Biology*, *34*(9), 1880-1892.e5. https://doi.org/10.1016/j.cub.2024.03.034

Kragel, P. A., Bianciardi, M., Hartley, L., Matthewson, G., Choi, J. K., Quigley, K. S., Wald, L. L., Wager, T. D., Barrett, L. F., & Satpute, A. B. (2019). Functional involvement of human periaqueductal gray and other midbrain nuclei in cognitive control. *Journal of Neuroscience*, *39*(31). https://doi.org/10.1523/JNEUROSCI.2043-18.2019

Krout, K. E., & Loewy, A. D. (2000). Periaqueductal gray matter projections to midline and intralaminar thalamic nuclei of the rat. *Journal of Comparative Neurology*, *424*(1). https://doi.org/10.1002/1096-9861(20000814)424:1<111::AID-CNE9>3.0.CO;2-3

Kuan, L., Li, Y., Lau, C., Feng, D., Bernard, A., Sunkin, S. M., Zeng, H., Dang, C., Hawrylycz, M., & Ng, L. (2015). Neuroinformatics of the allen mouse brain connectivity atlas. *Methods*, *73*. https://doi.org/10.1016/j.ymeth.2014.12.013

Lawen, A., Succi, I. K., Lichtman, D., Peterka, D. S., Abdus-Saboor, I. J., & Kahn, I. (2025). Dopamine release effects on striatal blood oxygenation and whole brain plasticity underlying associative learning. In *bioRxiv*. https://doi.org/10.1101/2025.07.20.665732

Lee, D., Seo, H., & Jung, M. W. (2012). Neural basis of reinforcement learning and decision making. In *Annual Review of Neuroscience* (Vol. 35). https://doi.org/10.1146/annurev-neuro-062111-150512

Lein, E. S., Hawrylycz, M. J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A. F., Boguski, M. S., Brockway, K. S., Byrnes, E. J., Chen, L., Chen, L., Chen, T. M., Chin, M. C., Chong, J., Crook, B. E., Czaplinska, A., Dang, C. N., Datta, S., … Jones, A. R. (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature*, *445*(7124), 168–176. https://doi.org/10.1038/nature05453

Levinson, M., Kolenda, J. P., Alexandrou, G. J., Escanilla, O., Cleland, T. A., Smith, D. M., & Linster, C. (2020). Context-Dependent Odor Learning Requires the Anterior Olfactory Nucleus. *Behavioral Neuroscience*, *134*(4). https://doi.org/10.1037/bne0000371

Lichtman, D., Bergmann, E., Kavushansky, A., Cohen, N., Levy, N. S., Levy, A. P., & Kahn, I. (2021). Structural and functional brain-wide alterations in A350V Iqsec2 mutant mice displaying autistic-like behavior. *Translational Psychiatry*, *11*(1). https://doi.org/10.1038/s41398-021-01289-8

Liska, A., Bertero, A., Gomolka, R., Sabbioni, M., Galbusera, A., Barsotti, N., Panzeri, S., Scattoni, M. L., Pasqualetti, M., & Gozzi, A. (2018). Homozygous loss of autism-risk gene cntnap2 results in reduced local and long-range prefrontal functional connectivity. *Cerebral Cortex*, *28*(4), 1141–1153. https://doi.org/10.1093/cercor/bhx022

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, *16*(5). https://doi.org/10.1523/jneurosci.16-05-01936.1996

Muller, T. H., Butler, J. L., Veselic, S., Miranda, B., Wallis, J. D., Dayan, P., Behrens, T. E. J., Kurth-Nelson, Z., & Kennerley, S. W. (2024). Distributional reinforcement learning in prefrontal cortex. *Nature Neuroscience*, *27*(3). https://doi.org/10.1038/s41593-023-01535-w

Nicholas, J., Amlang, C., Lin, C. Y. R., Montaser-Kouhsari, L., Desai, N., Pan, M. K., Kuo, S. H., & Shohamy, D. (2024). The Role of the Cerebellum in Learning to Predict Reward: Evidence from Cerebellar Ataxia. *Cerebellum*, *23*(4). https://doi.org/10.1007/s12311-023-01633-2

Nicholas, J., Daw, N. D., & Shohamy, D. (2022). Uncertainty alters the balance between incremental learning and episodic memory. *ELife*, *11*. https://doi.org/10.7554/ELIFE.81679

Nicola, S. M. (2007). The nucleus accumbens as part of a basal ganglia action selection circuit. In *Psychopharmacology* (Vol. 191, Issue 3). https://doi.org/10.1007/s00213-006-0510-4

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, *53*(3). https://doi.org/10.1016/j.jmp.2008.12.005

O'Doherty, J. P. (2004). Reward representations and reward-related learning in the human brain: Insights from neuroimaging. In *Current Opinion in Neurobiology* (Vol. 14, Issue 6). https://doi.org/10.1016/j.conb.2004.10.016

O'Doherty, J. P., Cockburn, J., & Pauli, W. M. (2017). Learning, Reward, and Decision Making. In *Annual Review of Psychology* (Vol. 68). https://doi.org/10.1146/annurev-psych-010416-044216

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron*, *38*(2). https://doi.org/10.1016/S0896-6273(03)00169-7

O'Doherty, J. P., Hampton, A., & Kim, H. (2007). Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of Sciences*, *1104*. https://doi.org/10.1196/annals.1390.022

Ozawa, T., Ycu, E. A., Kumar, A., Yeh, L. F., Ahmed, T., Koivumaa, J., & Johansen, J. P. (2017). A feedback neural circuit for calibrating aversive memory strength. *Nature Neuroscience*, *20*(1). https://doi.org/10.1038/nn.4439

Paton, J. J., Belova, M. A., Morrison, S. E., & Salzman, C. D. (2006). The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature*, *439*(7078). https://doi.org/10.1038/nature04490

R Core Team. (2024). R Core Team (2024). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria.* . http://www.R-project.org/

Reis, F. M. C. V., Lee, J. Y., Maesta-Pereira, S., Schuette, P. J., Chakerian, M., Liu, J., La-Vu, M. Q., Tobias, B. C., Ikebara, J. M., Kihara, A. H., Canteras, N. S., Kao, J. C., & Adhikari, A. (2021). Dorsal periaqueductal gray ensembles represent approach and avoidance states. *ELife*, *10*. https://doi.org/10.7554/eLife.64934

Rescorla, R. (1972). A Theory of Pavlovian Conditioning : Variations in the Effectiveness of Reinforcement and Nonreinforcement. *Classsical Conditioning II : Current Research and Theory*, 64–99. https://cir.nii.ac.jp/crid/1571698601305244544.bib?lang=en

Roy, M., Shohamy, D., Daw, N., Jepma, M., Wimmer, G. E., & Wager, T. D. (2014). Representation of aversive prediction errors in the human periaqueductal gray. *Nature Neuroscience*, *17*(11). https://doi.org/10.1038/nn.3832

RStudio Team. (2020). Integrated development environment for R. In *The Journal of Wildlife Management* (Vol. 75, Issue 8).

Salamone, J. D., & Correa, M. (2012). The Mysterious Motivational Functions of Mesolimbic Dopamine. In *Neuron* (Vol. 76, Issue 3). https://doi.org/10.1016/j.neuron.2012.10.021

Schoenbaum, G., Roesch, M. R., & Stalnaker, T. A. (2006). Orbitofrontal cortex, decision-making and drug addiction. In *Trends in Neurosciences* (Vol. 29, Issue 2). https://doi.org/10.1016/j.tins.2005.12.006

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306). https://doi.org/10.1126/science.275.5306.1593

Shusterman, R., Smear, M. C., Koulakov, A. A., & Rinberg, D. (2011). Precise olfactory responses tile the sniff cycle. *Nature Neuroscience*, *14*(8), 1039–1044. https://doi.org/10.1038/nn.2877

Sukikara, M. H., Mota-Ortiz, S. R., Baldo, M. V., Felício, L. F., & Canteras, N. S. (2006). A role for the periaqueductal gray in switching adaptive behavioral responses. *Journal of Neuroscience*, *26*(9). https://doi.org/10.1523/JNEUROSCI.4279-05.2006

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction Second edition, in progress*.

Tryon, V. L., & Mizumori, S. J. Y. (2018). A novel role for the periaqueductal gray in consummatory behavior. *Frontiers in Behavioral Neuroscience*, *12*. https://doi.org/10.3389/fnbeh.2018.00178 van Geen, C., & Gerraty, R. T. (2021). Hierarchical Bayesian models of reinforcement learning: Introduction and comparison to alternative methods. In *Journal of Mathematical Psychology* (Vol. 105). Academic Press Inc. https://doi.org/10.1016/j.jmp.2021.102602

Walker, R. A., Wright, K. M., Jhou, T. C., & McDannald, M. A. (2020). The ventrolateral periaqueductal grey updates fear via positive prediction error. *European Journal of Neuroscience*, *51*(3). https://doi.org/10.1111/ejn.14536

Winkelmeier, L., Filosa, C., Hartig, R., Scheller, M., Sack, M., Reinwald, J. R., Becker, R., Wolf, D., Gerchen, M. F., Sartorius, A., Meyer-Lindenberg, A., Weber-Fahr, W., Clemm von Hohenberg, C., Russo, E., & Kelsch, W. (2022). Striatal hub of dynamic and stabilized prediction coding in forebrain networks for olfactory reinforcement learning. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-30978-1

Wright, K. M., & McDannald, M. A. (2019). Ventrolateral periaqueductal gray neurons prioritize threat probability over fear output. *ELife*, *8*. https://doi.org/10.7554/eLife.45013