Article

Prefrontal to ventral tegmental area dynamics drive contingency degradation

https://doi.org/10.1038/s41586-026-10443-5

Received: 24 December 2024

Accepted: 24 March 2026

Published online: 06 May 2026

Check for updates

Madelyn M. Hjort[1,2,3], Zoe Q. Garrett[2,3,4], Adam G. Gordon[2,3], Ethan Ancell[2,5], Marta Trzeciak[2,4], Pei-Yun Lu[2,3], Michael R. Bruchas[2,3,4], Daniela M. Witten[2,5,6], Nicholas A. Steinmetz[2,7] & Garret D. Stuber[2,3,4] ✉

Cognitive flexibility refers to the adaptive neural processes that adjust learned behaviours as circumstances shift, supporting optimal decision-making and behavioural control. This includes the capacity to modify specific behaviours as the contingency between cues and rewards degrades. Across species[1–4], the medial prefrontal cortex (mPFC) has a well-established role in controlling contingency degradation[5]; however, the precise neural circuit mechanisms underlying this cognitive process remain unclear. To address this gap, we developed a quantitative model of cognitive flexibility that incorporates a meta-learning parameter into an established reward prediction error learning model[6,7]. Our meta-reward prediction error model significantly improves accurate representation of mouse cue-evoked licking behaviour in response to degraded or enhanced cue–reward associations. Using longitudinal two-photon calcium imaging and single-cell holographic optogenetics, we found that a subset of neurons in the mPFC specifically encode the contingency degradation in a significant and causal manner. Recognizing that behavioural flexibility probably requires interactions between the mPFC and canonical reward learning circuitry, we then examined how mPFC neural signalling during contingency degradation interacts with the ventral tegmental area (VTA)—a critical hub for reward processing[8]. Our imaging and optogenetics data show that mPFC sends this signal to VTA, with most mPFC→VTA neurons reflecting this transmission, and that selective optogenetic stimulation of these ensembles accelerates contingency degradation. These findings reveal how prefrontal circuits facilitate flexibility, selectively halting learned behaviours through connections with subcortical reward networks.

The ability to link valued outcomes with their predictive cues is a vital process that underlies decision-making for behavioural control[8]. It is equally essential to adapt behavioural strategies as cue-outcome relationships shift in dynamic landscapes rather than perseverating on an initially learned, but now outdated, strategy. Cognitive flexibility describes the process of modifying learned behaviours or strategies in changing environments, and is disrupted in many cognitive disorders including addiction[9]. Understanding how the brain selectively degrades the value of certain cue–reward associations while preserving others is crucial for developing future interventions for addiction, for which persistent high-value drug-cue associations can drive relapse[10]. Here we sought to develop a mechanistic understanding of how the brain supports learning to selectively stop certain behaviours while continuing others in the same context.

Successfully adapting behaviour after cue contingency changes hinges upon first detecting that the environment has changed, and then rapidly updating cue–reward values and behavioural choices accordingly. The prefrontal cortex (PFC) has been implicated heavily in this reversal learning process in humans[4,11] and other animals[1–3]. A sizeable body of research on prefrontal flexibility highlights the role of the lateral orbitofrontal cortex (OFC), as lesions in this region block behavioural adaptation after contingency reversals[5,12]. OFC signalling is highest immediately after contingency reversal[13,14], but often returns to baseline levels long before the full behavioural reversal takes place[14]. This activity timescale indicates that OFC may detect environmental change, but additional prefrontal circuits probably expedite cue revaluation and behavioural adaptation after contingency reversal.

Unlike the lateral OFC, lesions in the mPFC[15] decelerate behavioural updating after cue contingency changes[5,16–20] rather than abolishing reversal learning entirely. After mPFC lesions, animals exhibit perseverative responding to previously rewarded cues[17–19], indicating an inability to extinguish high-value contingences after repeated reward omissions. This suggests that the mPFC may contain neural circuitry that expedites behavioural updating in response to

[1]Graduate Program in Neuroscience, University of Washington, Seattle, WA, USA. [2]Center for Neurobiology of Addiction Pain and Emotion, University of Washington, Seattle, WA, USA. [3]Department of Anesthesiology and Pain Medicine, University of Washington, Seattle, WA, USA. [4]Department of Pharmacology, University of Washington, Seattle, WA, USA. [5]Department of Statistics, University of Washington, Seattle, WA, USA. [6]Department of Biostatistics, University of Washington, Seattle, WA, USA. [7]Department of Neurobiology and Biophysics, University of Washington, Seattle, WA, USA. ✉e-mail: gstuber@uw.edu

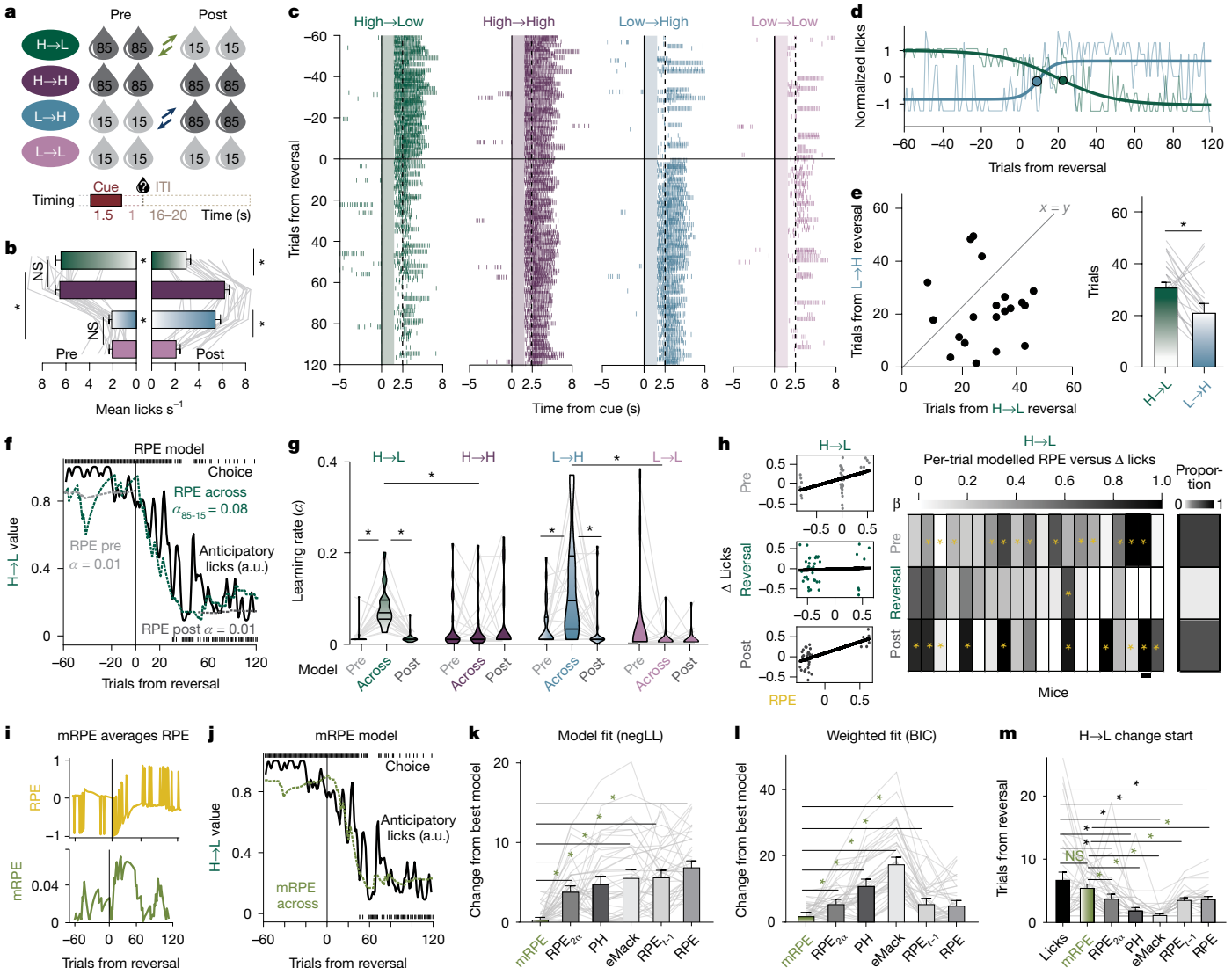


Fig. 1 | Meta-RPE model explains reversal learning behaviour. a, Behaviour task schematic. Animals experience a single reversal after stable mastery of initial Pavlovian contingencies. **b**, Anticipatory licking between cue and reward delivery quantifies task performance (licks × cue type $P < 0.0001$; $n = 20$ mice). **c**, Licking raster plot from an example mouse across reversal. **d**, Fits of anticipatory licking curves display different inflection points for contingency degradation (H→L) and enhancement (L→H) in an example mouse. **e**, Scatter plot of behavioural inflection (left) indicates that most mice change L→H licking before H→L, represented at the population level (right) with a bar graph (H→L > L→H, $P = 0.0273$; $n = 20$ mice). **f**, Sample RPE-based value curves for licking behaviour pre-, post- or across reversal. **g**, Quantification of RPE gain (learning rate) for models from **f** in $n = 20$ mice (cue × time $P = 0.0008$). **h**, Correlations between trial-modelled RPE and behavioural change on the next trial at the individual (left) and population (right) level are not significant in most mice during the reversal ($P = 0.3299$; $n = 20$ mice). **i**, Visualization of the relationship between RPE and mRPE. **j**, mRPE model fit to the same data as in **f**. **k**, Comparison of model fit on $n = 40$ mice to licking behaviour using an unweighted metric (negLL) between the meta-RPE model (mRPE), RPE models and other models with dynamic learning rate including the RPE model with split learning rate for positive and negative errors ($RPE_{2\alpha}$), Pearce-Hall (PH), extended Mackintosh (eMack) and RPE model with learning rate modulated by the previous RPE (RPE_{t-1}). The mRPE model fits significantly better than the other models ($P < 0.0001$ for all comparisons; $n = 40$ mice). **l**, As in **k**, but with a weighted fit metric, Bayes information criterion (BIC), that penalizes models for additional free parameters. The mRPE model still fits significantly better than the other models with this penalty ($n = 40$ mice; $RPE_{2\alpha}$, $P = 0.0064$; PH, $P < 0.0001$; eMack, $P < 0.0001$; RPE_{t-1}, $P = 0.0258$; RPE, $P = 0.0339$). **m**, Comparison of the trial at the point at which value degradation begins (Methods) for all models, in addition to anticipatory licking behaviour. Only the mRPE model has a non-significant difference from the behaviour ($n = 40$ mice, mRPE, $P = 0.2086$; $RPE_{2\alpha}$, $P = 0.0255$; PH, $P < 0.0001$; eMack, $P < 0.0001$; RPE_{t-1}, $P = 0.0002$; RPE, $P = 0.0002$). Error bars denote ± s.e.m. *$P < 0.05$. See Supplementary Table 1 for more statistical information, including more post hoc comparisons between models (Extended Data Fig. 1), sidedness and corrections for multiple comparisons. NS, non-significant.

contingency changes. Here we develop a quantitative model to simulate this flexibility, using task variables that can be related to mPFC activity at single-cell resolution. Our longitudinal in vivo two-photon calcium imaging demonstrates that a subset of neurons in mPFC causally encode a selective contingency degradation error signal that is transmitted to the VTA—a critical subcortical learning hub that drives motivated behaviour.

To isolate a quantitative behavioural flexibility signal, we trained head-fixed mice on a Pavlovian reversal learning task (Fig. 1a) in which four unique odour cues initially signalled reward deliveries with either 85% or 15% probability (two cues each). After stable mastery of the initial Pavlovian contingencies, mice experienced a single contingency reversal, for which one cue–reward association reversed from high to low value (contingency degradation; H→L trials). In the same reversal

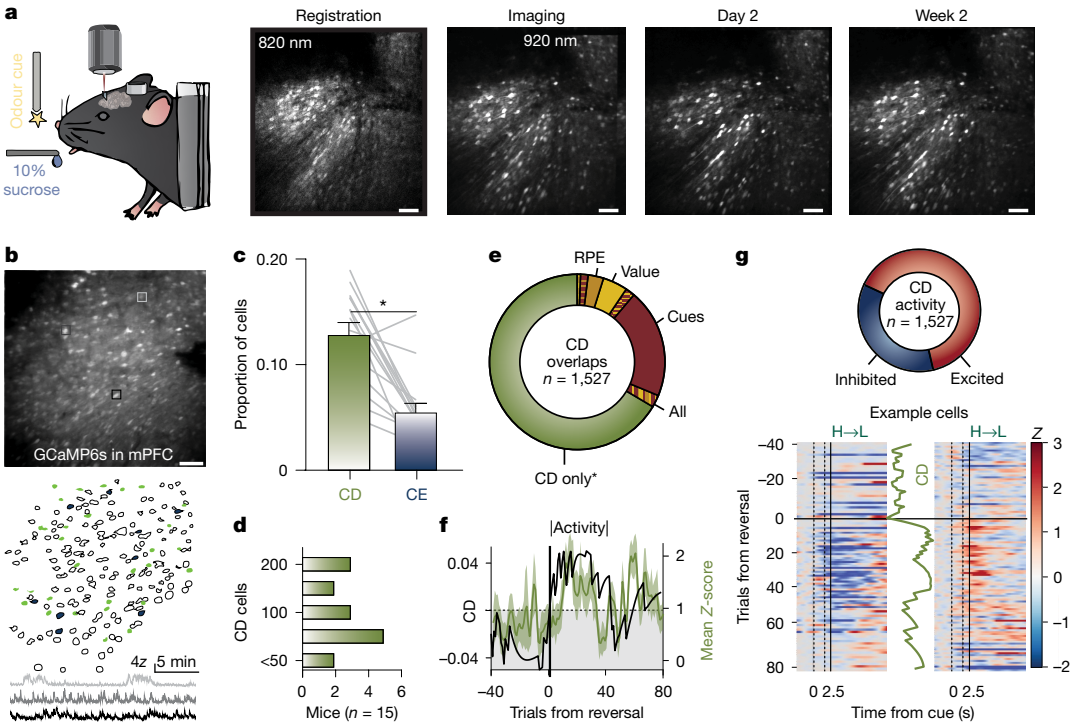


Fig. 2 | A neural correlate for CD in mPFC. a, Schematic of head-fixed two-photon calcium imaging strategy, with single neurons tracked over days and weeks. **b**, Example FOV with extracted transients and GLM-isolated (Extended Data Fig. 2 and Methods) CD (green) and CE (blue) cells. **c**, Proportions of CD and CE cells in mPFC ($P < 0.0001$, $n = 15$ mice). **d**, Distribution of CD cell numbers across the dataset ($n = 1,527$ out of 11,792 cells from $n = 15$ mice). **e**, Overlap comparison for all CD cells with other GLM categories. There are significantly more cells that do not overlap than any other comparison ($P < 0.0001$, $n = 15$ mice). **f**,**g**, Average trial amplitude (**f**) of CD cells in an example mouse (green) parallels the CD signal (black), with a combination of excited and inhibited cells (**g**). See Supplementary Table 1 for more statistical information, including sidedness and corrections for multiple comparisons. See Extended Data Fig. 3 for implant placements. *$P < 0.05$. Error bars denote ± s.e.m. Scale bars, 100 µm.

session, another cue–reward association was enhanced (L→H trials), whereas two value-matched control cues did not reverse (H→H, L→L). Licking between cue presentation and reward delivery (0–2.5 s), reflecting the animal's anticipation of a future reward[21,22], was used to quantify task performance and was not significantly different between matched contingency cues before, after or across stable days after reversal (Fig. 1b). The contingency reversal was not indicated to the mice, so anticipatory licking towards reversed cues took many trials to change and several sessions to fully re-stabilize (Fig. 1c). Notably, the midpoints of the H→L degradation and L→H enhancement behaviours were distinct and dissociable (Fig. 1d), with the response to degradation taking longer than enhancement on average (Fig. 1e) even when the overall sequence of trials is considered (Extended Data Fig. 1a). This result indicates that animals commit several trial errors using their original strategy before they change their behaviour, and that learning from cumulative positive errors and cumulative negative errors occurs at different rates.

On the basis of our behavioural observations, we designed a computational model of the learning process in this task. Traditional reinforcement learning models[7] learn by updating a value function using reward prediction errors (RPEs)[23]—the difference between experienced reward and what had been predicted at a static gain (learning rate[7,23]). The Rescorla–Wagner RPE model[6,7], is able to capture a contingency degradation trajectory (Fig. 1f) but displays notable differences from animal behaviour during our reversal learning task. First, model fits that include reversal behaviour display increased learning rates compared with unreversed cues or stable behaviour days (Fig. 1g). Although RPE predicts next-trial behavioural change before and after the reversal, there is no significant linear relationship between RPE and behavioural change during the reversal period in 19 out of 20 mice (Fig. 1h). This discrepancy is attributable to a lack of behavioural change immediately after the reversal when RPEs are highest (Fig. 1h), with change instead at intermediate RPEs. These observations indicate that animals increase the weight of mid-reversal RPEs (compared with early or late trials) in the behavioural implementation of the contingency reversal.

The lag between task and behavioural reversal points suggests that animals may accrue errors over several trials before they change their behaviour, requiring an integration of RPEs. This integration can take the form of another error signal that computes the average RPE over several trials. Such a meta-RPE (mRPE) signal is highest mid-reversal when behaviour is rapidly changing and can be calculated online from the rolling average of individual RPEs (Fig. 1i). Allowing mRPE to modulate RPE gain (equations (3)–(6); Methods) produces a contingency degradation curve (Fig. 1j) that fits significantly better than single-trial error (RPE) models and other established models with dynamic learning rates (Extended Data Fig. 1b–e) when considering the animal's choice (Fig. 1k,l) or anticipatory licking (Fig. 1m) behaviour (examined in $n = 40$ mice). Incorporating different integration rates for contingency degradation and enhancement recapitulates the delayed behavioural change onset and differences in their temporal dynamics. As the mRPE is high during flexible (reversal) behaviour, mRPE could represent a quantitative cognitive flexibility signal with dissociable and calculable signals for contingency degradation (CD) and contingency enhancement (CE). The mRPE model, behaviour fitting dataset and figure source data used for hypothesis testing are available at github.com/stuberlab/Hjort-et-al.-2026-PFC-and-reversal-learning and ref. 24.

mPFC lesions selectively impair learning from concentrated negative errors[17–19], thus mPFC neurons could be a substrate of CD. Supporting this possibility, standard reinforcement learning models that lack a CD term fail to accurately predict the effects of mPFC lesions on contingency degradation[20]. We performed longitudinal two-photon calcium imaging of GCaMP6s[25] through microprisms[26,27] to image 11,792 mPFC neurons across reversal learning (Fig. 2a) and used a modified

generalized linear model (GLM)[28] to test for significant encoding of task and cognitive variables for each neuron (Extended Data Fig. 2 and Methods) (package available at github.com/stuberlab). A subpopulation (13.1 ± 1.1%) of mPFC neurons (Fig. 2b) displayed significant CD encoding (Fig. 2c,d) that was largely distinct from other task variables including cue delivery and trial value (Fig. 2e). The activity of individual excited (983 out of 1,527, 64.3 ± 4.3%) and inhibited (544 out of 1,527, 35.6 ± 4.3%) CD cells (Fig. 2g), along with their average activity (Fig. 2f) parallels the modelled CD signal. The neural CD encoding is time-locked to contingency degradation trials, as shifting the CD signal within (Extended Data Fig. 2 and Methods) or across (Extended Data Fig. 4j) days did not recapitulate the proportion of significant CD cells. Population encoding of cue meaning[21] (Extended Data Fig. 4a–i) and CD (Extended Data Fig. 4k) was largely stable across reversal and did not decay with time, rendering a representational drift-driven explanation unlikely. Altogether, a subset of mPFC cells act as a neural substrate for a CD signal and may drive behavioural flexibility during contingency degradation.

GLM modelling can only correlatively link CD and mPFC activity. To overcome this, we used holographic single-cell optogenetics using a spatial light modulator (SLM)[29] to test this relationship causally by selectively activating CD encoding neurons in the mPFC and quantifying subsequent effects on recorded and modelled value-guided licking behaviour. Mice expressing both the two-photon compatible excitatory opsin ChRmine[30] and GCaMP6s[25] in mPFC (Fig. 3a), performed the reversal task to first identify CD neurons (Fig. 3b) and trained subsequently to stable post-reversal behaviour. As we hypothesized that CD ensemble activity decreases value-driven licking behaviour, the CD ensemble received selective optogenetic holographic activation[30] (Fig. 3c) on half of trials during presentation of the unreversed high-value cue (H→H), whereas the other half of trials served as controls (Fig. 3d and Methods). Patterned optogenetic stimulation of neuronal ensembles of CD (76 ± 12 cells per session) or a matched number of control cells produced robust activation in the targeted neurons in the field of view (FOV), validating the spatial and temporal specificity of this approach (Fig. 3c). In SLM experiment 1, CD cell activation occurred on half of trials intermixed randomly with unstimulated trials (Fig. 3d). There were significantly fewer anticipatory licks on CD stimulated trials compared with the trials with no stimulation (Fig. 3e,f). In SLM experiment 2, CD activation trials occurred randomly interleaved with a size-matched control ensemble of neurons that did not significantly encode any task variables (Fig. 3d). As in SLM experiment 1, there were significantly fewer anticipatory licks on CD stimulated trials compared with control trials in SLM experiment 2 (Fig. 3g,h). In SLM experiment 3, CD activation also occurred on 50% of trials randomly intermingled with a size-matched control ensemble of cells that encoded another GLM variable (cue) but not CD (Fig. 3d). These mice also displayed a significant reduction in anticipatory licking during CD cell stimulation (Fig. 3i,j). There was no significant difference in anticipatory licking to control stimulation across the three experiments (Fig. 3k), suggesting effects from CD cell stimulation primarily drove the observed behavioural effects. As CD cell activations were much larger in all three experiments compared with activity levels during the behavioural tasks (Fig. 2g versus Fig. 3c), mice could be expected to have a much larger amplitude CD signal in the behavioural model. Fitting the mRPE model to behavioural data from all SLM experiments supports this conclusion, with significantly higher model-fit CD amplitudes on the stimulated high-value cue compared with the unstimulated high-value cue (Fig. 3l). Thus, SLM stimulation of identified CD neurons reduced cue-evoked licking, suggesting that mPFC neurons causally encode CD to drive contingency degradation.

The mRPE model calculates mRPE from a rolling average of single-trial RPEs. Consequently, mPFC cells need RPE information to calculate CD. Given the well-established role of VTA dopamine neurons in RPE coding[23,31] and strong reciprocal mPFC and VTA connectivity[32,33], the mPFC CD cells could receive this RPE information from the VTA. Thus, we used fibre photometry to record changes in the dopamine indicator GRAB-DA3h[34] (Fig. 4a), reflecting VTA→mPFC dopamine release during contingency reversal. Consistent with previous reports[35], VTA→mPFC dopamine increased during the reversal (Fig. 4b,c), and this was largely attributable to signalling during the contingency enhancement (L→H), supporting that unexpected RPE coding (Fig. 4d) may be one of the many functions[36] of dopamine signalling in the mPFC. GLM analysis of the dopamine signal revealed significant cue, value, CE and RPE encoding (Fig. 4e). These results indicate that mPFC receives RPE information from VTA, as is required to calculate CD for contingency degradation.

The mRPE model degrades contingencies when a rising CD signal increases the gain of omission-related RPEs. Accordingly, the mRPE model predicts that CD should influence single-trial RPE signalling. Past neuroanatomical findings are consistent with this possibility as VTA dopamine neurons are the canonical substrate of RPE[23,31], and approximately 10% of mPFC cells project directly to VTA[32,37]. Fibre photometry-mediated mPFC→VTA terminal calcium recordings during reversal (Fig. 4p) displayed significant CD encoding (Fig. 4q), indicating that the CD signal directly reaches the neural substrate of RPE (VTA). Single-cell calcium imaging of VTA-projecting mPFC cells (Fig. 4f) revealed that most (61.5 ± 9.45%) significantly encode CD signalling (Fig. 4g,h)—a significant increase in proportion compared with the total mPFC population (Figs. 2c and 4i,). Together, these results substantiate that mPFC→VTA circuitry preferentially represents a CD signal for contingency degradation (Fig. 4j).

If VTA→mPFC dopamine RPE signalling is necessary for mPFC to compute mRPE and drive the H→L reversal, then infusing dopamine antagonists into the mPFC (Fig. 4k) should impair CD signalling (Fig. 4l) and, subsequently, the time course of reversal. Consistent with this prediction, infusing a cocktail of D1R and D2R antagonists (SCH23390 (SCH) and raclopride (RAC)) into mPFC during reversal (Fig. 4m,n) reduces CD encoding in the calcium activity of mPFC→VTA neuron terminals (Fig. 4q), CD amplitude in the mRPE model (Fig. 4o) and the reversal midpoint (Fig. 4o) compared with saline controls. These results provide causal evidence that reciprocal mPFC and VTA signalling drives contingency degradation during flexible behaviour.

The mRPE model implements CD by increasing the gain of repeated negative RPEs. Traditionally, the VTA represents negative RPE through reduced activity of dopaminergic (VTA_{DA}) neurons[23,31,38], or through increased activity of the GABAergic (VTA_{GABA}) neurons that inhibit VTA_{DA} neurons[39,40]. Thus, mPFC→VTA cells encoding CD should interact with either target to enhance negative single-trial RPEs and drive CD behaviour. Simultaneous interaction with both populations could also explain the split excitation or inhibition profile of CD cells in mPFC (Figs. 2g and 4h). Imaging the ex vivo calcium activity of VTA_{GABA} and VTA_{DA} neurons in brain slices with concurrent mPFC terminal stimulation (Fig. 5a) supports this hypothesis, with 50.1 ± 2.7% of the 581 VTA_{GABA} neurons responding (Fig. 5b,f), primarily through excitation (Fig. 5c,g), alongside significantly fewer VTA_{DA} neurons (Fig. 5d,f, 35.8 ± 2.6% of n = 1,084 cells) split evenly between excitation and inhibition (Fig. 5e,g). This heterogeneity in VTA_{DA} response could explain the mixed activity of mPFC CD cells (Figs. 2g and 4h). Nevertheless, because the strongest modulated population was the VTA_{GABA} neurons that inhibit VTA_{DA} neurons[39–41] during negative RPEs, exciting the mPFC→VTA projection should drive contingency degradation behaviour.

To test in vivo the hypothesis that exciting mPFC→VTA projections should causally expedite contingency degradation during reversal learning, we injected virus to express channelrhodopsin in mPFC neurons and stimulated through optical fibres placed over the VTA while mice performed the reversal task. Optogenetic activation of the mPFC→VTA projection outside of the task context produced neither rewarding nor aversive behavioural phenotypes in a real-time preference test (Fig. 5i), supporting the idea that this circuit modulates learning and behavioural flexibility rather than innate approach or avoidance. Within the reversal learning task, stimulating the mPFC→VTA

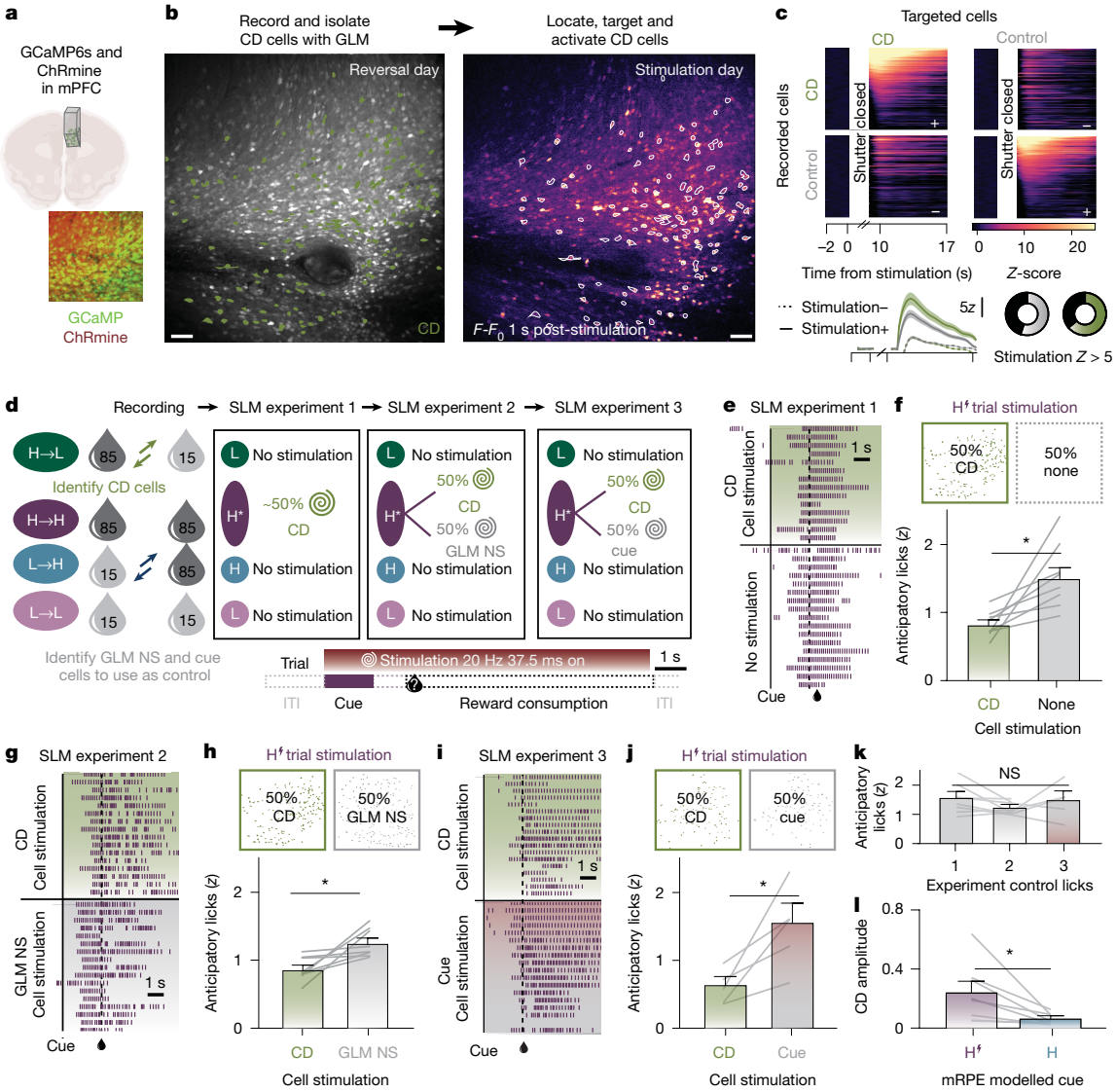


Fig. 3 | Single-cell holographic optogenetic activation of mPFC CD cells causes CD. a, Schematic showing expression of both GCaMP6s and ChRmine in mPFC. **b**, Imaging mPFC cells expressing GCaMP6s and ChRmine during the reversal task isolated CD cells with GLM. The CD ensemble (left) was subsequently activated after post-reversal stability using holographic optogenetics (SLM) (right). **c**, Targeted activation of cells in an example mouse. Note higher levels of activation in most cells when cells are targeted. **d**, Schematic of SLM experiments. Spiral indicates SLM stimulation. The H→H cue experienced stimulation (H′), whereas the L→H served as an unstimulated value-matched control (H). ITI, inter-trial interval. **e**, Example licking behaviour for SLM experiment 1 in which CD cells were stimulated (top) in half of trials. Trials are sequential within a group. **f**, There were significantly fewer anticipatory licks during CD stimulation in SLM experiment 1 ($P = 0.0054$, $n = 8$ mice). **g**, Example licking behaviour during CD (top) or GLM non-stimulated control (bottom) ensemble stimulation in SLM experiment 2. **h**, There were significantly fewer licks on CD stimulation trials compared with GLM–control trials in SLM experiment 2 ($P = 0.0077$, $n = 8$ mice). **i**, Example licking behaviour during CD (top) or cue cell control (bottom) ensemble stimulation in SLM experiment 3 **j**, There were significantly fewer licks on CD stimulation trials compared with cue cell control trials in SLM experiment 3 ($P = 0.0139$, $n = 5$ mice). **k**, There was no significant difference in number of licks to control stimulation across the three experiments (P = 0.3460, n = 8 mice). l, Larger mRPE-modelled CD amplitude for H′ compared with H trials suggests CD cell photostimulation produced a larger CD signal ($P = 0.0250$, $n = 8$ mice). See Supplementary Table 1 for more statistical information, including more post hoc comparisons, sidedness and corrections for multiple comparisons. See Extended Data Fig. 3 for implant placements. Error bars denote ± s.e.m. *$P < 0.05$. Scale bars, 100 µm.

projection (Fig. 5j,k) after the mice had acquired the initial Pavlovian association but before the reversal generated a modest but significant reduction in anticipatory licking (Fig. 5k–m), consistent with our SLM results (Fig. 3). By contrast, optogenetic stimulation of mPFC→VTA activity concurrent with contingency degradation significantly (Fig. 5m), and near instantly (Fig. 5n), reduced anticipatory licking towards the H→L cue. mRPE-mediated value models of the reversal stimulation behaviour display significantly enhanced CD amplitudes compared with mCherry expressing control mice (Fig. 5o). Chemogenetic inhibition of mPFC→VTA neurons significantly decreased CD amplitude and slowed reversal (Extended Data Fig. 5). Together, these causal experiments present further evidence that mPFC→VTA signalling drives contingency degradation.

It is possible that these effects, and those from Fig. 3, arose from a transient suppression of behavioural responding with CD cell stimulation instead of a true learned change. To test whether the anticipatory lick reduction outlasted the stimulation (supporting the learning hypothesis) or reverted to pre-reversal levels (supporting the transient suppression hypothesis), we compared post-reversal sessions in ChR2 mice with and without mPFC→VTA stimulation. Consistent

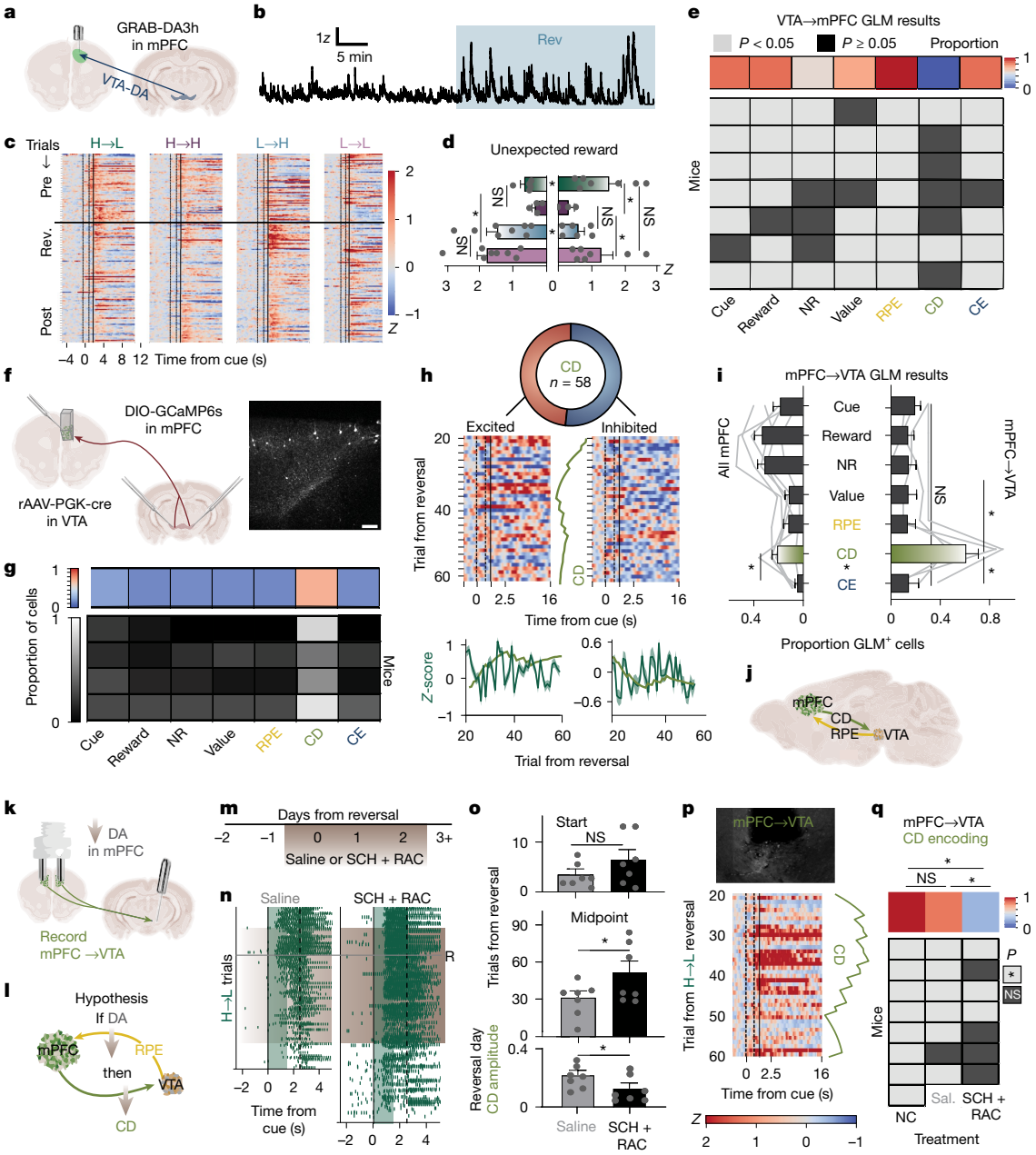


Fig. 4 | mPFC→VTA circuit implements mRPE model. a–**c**, mPFC dopamine photometry recordings (**a**) displayed an amplitude increase (**b**) after contingency reversal (**c**). **d**, Mean dopamine amplitude after unanticipated rewards (rewards with no preceding licks, $P < 0.0001$, $n = 8$ mice). **e**, GLM results suggest dopamine encodes RPE in mPFC. **f**, mPFC→VTA two-photon with example FOV. **g**, mPFC→VTA single-cell GLM results with proportion of cells (Extended Data Fig. 4l) from $n = 4$ individual mice (centre) and pooled average (top). **h**, mPFC→VTA CD cell activity (top) with individual examples (centre) and averages (bottom). **i**, Comparison of all encoding in mPFC ($n = 6$ mice from Fig. 2) and mPFC→VTA ($n = 4$ mice) cells. There are significantly more CD cells in the mPFC→VTA dataset ($P < 0.0001$). **j**, Circuit diagram representing results from **e** and **i**. **k**, Implant of bilateral cannula facilitates dopamine pharmacology during reversal alongside calcium recordings of mPFC terminals. **l**–**q**, This experiment tests the hypothesis (**l**) that impairing dopaminergic input to mPFC during reversal (**m**) will impair CD signalling as dopamine in mPFC conveys RPE (**e**) and the mRPE model suggests CD computations involve averaging RPEs. Lick data from sample mice infused with saline (grey) or SCH + RAC illustrates impaired H→L reversal (R) with treatment (**n**), quantified (**o**) as a significantly delayed midpoint in the treatment group ($P = 0.0368$, $n = 7$ mice per group) and smaller CD amplitude on the reversal day ($P = 0.0394$, $n = 7$ mice per group). In untreated mice, sample mPFC→VTA terminals encode CD (**p**), as does the population (**q**). This effect is reduced significantly with SCH + RAC infusion across reversal ($P = 0.0058$ untreated versus SCH + RAC, $P = 0.0445$ saline versus SCH + RAC, $n = 8$ mice (untreated), 7 (saline) and 7 (SCH + RAC). See Supplementary Table 1 for more statistical information, including more post hoc comparisons, sidedness and corrections for multiple comparisons. See Extended Data Fig. 3 for implant placements. NC, no cannula implanted. *$P < 0.05$. Error bars denote ± s.e.m. Scale bar, 100 µm.

with the learning hypothesis, licking remains reduced significantly from pre-reversal levels, and comparable with post-reversal levels in control mice (Fig. 5m). Therefore, the mPFC→VTA projection causally implements a CD signal to drive contingency degradation learning (Fig. 5p).

Our work indicates that behavioural change occurs more rapidly during contingency reversal because mRPE signals increase the gain of individual RPEs. These data provide an explanation for the long-standing observation that learning rates are expedited during reward contingency changes[42]. Our quantitative modelling suggests that both

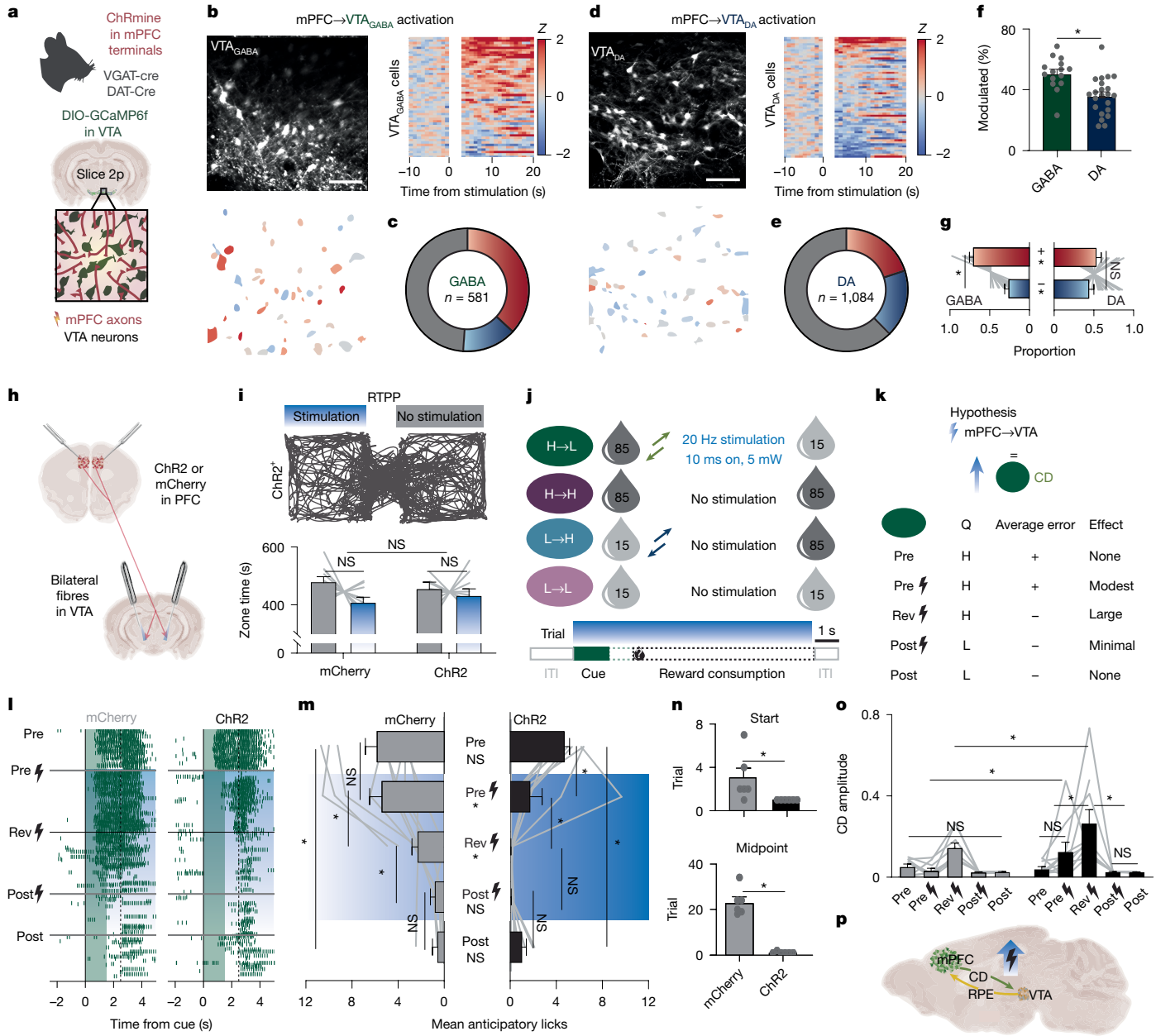


Fig. 5 | Activating mPFC→VTA circuitry increases CD and expedites contingency degradation during reversal learning. a, Slice 2p experiment in which VGAT-Cre (GABA cohort) or DAT-Cre (dopamine cohort) mice expressing ChRmine in mPFC and DIO-GCaMP6f in VTA received intermittent full-field stimulation. **b**, Sample VTA$_{GABA}$ FOV with mPFC stimulation aligned activity (right) and average region of interest (ROI) activity map below. **c**, Population statistics across 581 cells from 16 slices and three mice comparing excited (red), inhibited (blue) and no (grey) response. Cells were classified using one-sample *t*-test compared with a test population mean of zero. **d**, Sample VTA$_{DA}$ FOV and results, as in **b**. **e**. Population statistics, as in **c**, across 1,084 VTA$_{DA}$ cells from 22 slices and three mice. **f**,**g**, There are significantly more modulated VTA$_{GABA}$ than VTA$_{DA}$ cells ($P = 0.0007$) (**f**), with more excited VTA$_{GABA}$ than VTA$_{DA}$ cells ($P = 0.0227$) overall (**g**). **h**, Experimental strategy schematic. **i**, mPFC→VTA real-time place preference does not have significant behavioural effects ($P = 0.8725$, $n = 10$ mice per group). **j**, Task stimulation strategy. After stable acquisition of the initial cue contingencies, mice experienced 1 day of stimulation concurrent with stable cue contingencies, and then subsequent stimulation as contingencies reversed within the task. **k**, Hypothesis of mPFC→VTA stimulation effects based on CD prevalence in each condition. **l**, Lick raster plots (as in Fig. 1c) before, during and after stimulation for example mice. **m**, There were significantly fewer anticipatory licks between ChR2$^+$ (right) and control (left) mice after contingency reversal, indicating faster degradation ($P = 0.0167$, $n = 10$ mice per group). The evoked behaviour change persists beyond the active stimulation period, with no significant difference in anticipatory licks between control and ChR2 groups post reversal ($P = 0.9198$, $n = 6$ mice per group) or for ChR2 mice on stimulation and no stimulation days post reversal ($P = 0.4377$, $n = 6$ mice). **n**, Both the start (top) and midpoint (bottom) of reversal occur significantly early in the ChR2 group ($P = 0.0493$ and $P < 0.0001$, respectively, $n = 6$ mice per group) **o**, Stimulation concurrent with reversal significantly increases modelled CD amplitude (ChR2 pre+ versus rev+ $P = 0.0154$, $n = 10$ mice), post+ versus rev+ $P < 0.0001$, $n = 6$ mice), mCherry versus ChR2 rev+ $P = 0.0058$ ($n = 10$ mice per group) **p**, Circuit schematic summarizing results, as in Fig. 4j. See Supplementary Table 1 for more statistical information, including more post hoc comparisons, sidedness and corrections for multiple comparisons. See Extended Data Fig. 3 for implant placements. RTPP, real-time place preference. Error bars denote ± s.e.m. *$P < 0.05$. Scale bar, 50 µm.

single-trial errors (RPEs) and their multi-trial averages (mRPEs) expedite learning when environmental contingencies fluctuate (Fig. 1). Other groups have developed meta-learning models that show variable tuning of learning rates in rodent models[13,43–48] through various assumptions (Extended Data Fig. 1), but they do not outperform the mRPE model in our task (Fig. 1). These earlier models predict that the largest

behavioural changes should coincide with the biggest single-trial errors. This is incongruent with the behaviour we observed, as animals wait for several trials to change their anticipatory licking, rendering the largest behavioural change at intermediate RPEs (Fig. 1). Our meta-RPE model includes a meta-learning signal (mRPE) that peaks at an intermediate point in the reversal, significantly improving value function fits to behaviour (Fig. 1). The idea of a meta-learning signal peaking mid-reversal is consistent with a human PFC neuroimaging study from Behrens and colleagues[49], who found a 'volatility signal' that peaked mid-reversal. Our work advances this idea by accounting for the difference between learning from positive versus negative errors, which are distinct. In our study, dopamine release in mPFC (Fig. 4) significantly encodes positive CE, whereas negative CD is reflected in the activity within mPFC→VTA projection neurons (Figs. 4 and 5). Collectively, this suggests that distinct neural circuit elements contribute uniquely to CE versus CD needed for cognitive flexibility.

The previously reported mPFC lesion effects on contingency reversal[5,16–20] may be explained partly by the loss of CD cell activity, which aligns with observed changes in neural dynamics. Several other groups have reported altered mPFC firing patterns during contingency changes, including Bissonnette and colleagues'[50] conflict cells, Malagon-Vina and colleagues'[51] low behavioural performance cells, Karlsson and colleagues'[52] increased signalling volatility, Powell and colleagues'[53] lower correlation and Rich and colleagues'[54] cells tracking proportion incorrect. Considered individually, increased CD cell activity only during reversal would increase population variability ('volatility' and 'lower correlation') that is otherwise fairly stable before[21] and after reversal (Extended Data Fig. 4). In our task, the CD cells change their activity during the 'conflict' between old and new contingency expectations because, as mice choose to lick in anticipation to the H→L cue, subsequent reward delivery is unlikely (15%), in 'conflict' with the pre-reversal 85% probability. As animals make continued seeking errors towards the devalued cue, they are making 'incorrect choices' impairing 'behavioural performance'. Therefore, it is likely that the activity of the mPFC CD ensemble is at least partially responsible for previously observed lesion effects and mPFC firing changes during reversal.

Quantitative identification of CD encoding in mPFC→VTA circuitry (Fig. 4) and causally linking this to contingency degradation behaviour (Figs. 4 and 5 and Extended Data Fig. 5) establishes a clear answer regarding how prefrontal (mPFC) and subcortical (VTA) areas coordinate changes in learned behaviours during cognitive flexibility. This reciprocal interaction involves RPE signals in VTA→mPFC dopamine (Fig. 4) that mPFC→VTA cells use to compute a CD signal (Fig. 4), which is sent back to VTA_{GABA} and VTA_{DA} neurons (Fig. 5). Overall, PFC terminal stimulation drove primarily excitatory responses in VTA neurons, largely in the VTA_{GABA} population with more heterogeneous responses in VTA_{DA} neurons. This functional connectivity profile aligns with previous cell-type-specific retrograde tracing studies of VTA neurons[37,55,56]. Previous work by Spellman and colleagues[27] has indirectly suggested a role for mPFC→VTA circuitry in cognitive flexibility. In their work, post-reward firing changes during reversal were not restricted preferentially to thalamic or striatal-projecting mPFC neurons, but localized spatially to deep layer V projection neurons[27], notably where mPFC→VTA neurons are concentrated[32,37]. Furthermore, patterned optogenetic stimulation of CD neurons expedites contingency degradation (Fig. 3) providing a causal link between mPFC ensemble activity and cognitive flexibility.

Our work establishes quantitative CE and CD signals critical for understanding neural circuits for cognitive flexibility. Although we demonstrate that some mPFC neurons and mPFC→VTA neurons in particular drive contingency degradation, these computational properties may also be reflected in many brain regions important for motivated behaviours, which interface either directly or indirectly with the mPFC and VTA. It remains to be seen whether these circuit mechanisms are impaired in addiction models, although we would speculate CD encoding cells may be compromised in some form after extended volitional access to drugs of abuse. Specifically, hypofrontality associated with active substance use disorder[9,57–61] could include loss of function in the mPFC→VTA pathway. This could manifest in the reported deficits in cognitive control and flexibility related to drug-associated cues[9,57–61].

This work provides a quantitative, mechanistic explanation for how certain prefrontal circuits permit flexibility to selectively stop learned behaviours through interactions with core subcortical reward systems.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-026-10443-5.

1. Kesner, R. P. & Churchwell, J. C. An analysis of rat prefrontal cortex in mediating executive function. *Neurobiol. Learn. Mem.* **96**, 417–431 (2011).
2. Dalley, J. W., Cardinal, R. N. & Robbins, T. W. Prefrontal executive and cognitive functions in rodents: neural and neurochemical substrates. *Neurosci. Biobehav. Rev.* **28**, 771–784 (2004).
3. Güntürkün, O. The avian 'prefrontal cortex' and cognition. *Curr. Opin. Neurobiol.* **15**, 686–693 (2005).
4. Milner, B. Effects of different brain lesions on card sorting: the role of the frontal lobes. *Arch. Neurol.* **9**, 90–100 (1963).
5. Izquierdo, A., Brigman, J. L., Radke, A. K., Rudebeck, P. H. & Holmes, A. The neural basis of reversal learning: an updated perspective. *Neuroscience* **345**, 12–26 (2017).
6. Rescorla, R. & Wagner, A. in *Classical Conditioning II: Current Research and Theory* (eds Black, A. & Prokosy, W.) 64–99 (Appleton-Century-Crofts, 1972).
7. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* (MIT Press, 2018).
8. Stuber, G. D. Neurocircuits for motivation. *Science* **382**, 394–398 (2023).
9. Faustino, B., Oliveira, J. & Lopes, P. Diagnostic precision of the Wisconsin Card Sorting Test in assessing cognitive deficits in substance use disorders. *Appl. Neuropsychol. Adult* **28**, 165–172 (2021).
10. Sinha, R. & Li, C. S. R. Imaging stress- and cue-induced drug and alcohol craving: association with relapse and clinical implications. *Drug Alcohol Rev.* **26**, 25–31 (2007).
11. Kim, C., Johnson, N. F., Cilles, S. E. & Gold, B. T. Common and distinct mechanisms of cognitive flexibility in prefrontal cortex. *J. Neurosci.* **31**, 4771–4779 (2011).
12. Ogg, M. C. et al. Locus coeruleus norepinephrine neurons facilitate orbitofrontal cortex remapping and behavioral flexibility. *Cell Rep.* **44**, 116687 (2025).
13. Namboodiri, V. M. K. et al. Relative salience signaling within a thalamo-orbitofrontal circuit governs learning rate. *Curr. Biol.* **31**, 5176–5191 (2021).
14. Banerjee, A. et al. Value-guided remapping of sensory cortex by lateral orbitofrontal cortex. *Nature* **585**, 245–250 (2020).
15. Laubach, M., Amarante, L. M., Swanson, K. & White, S. R. What, if anything, is rodent prefrontal cortex? *eNeuro* **5**, ENEURO.0315-18.2018 (2018).
16. Bussey, T. J., Muir, J. L., Everitt, B. J. & Robbins, T. W. Triple dissociation of anterior cingulate, posterior cingulate, and medial frontal cortices on visual discrimination tasks using a touchscreen testing procedure for the rat. *Behav. Neurosci.* **111**, 920–936 (1997).
17. Biró, S., Lasztóczi, B. & Klausberger, T. A visual two-choice rule-switch task for head-fixed mice. *Front. Behav. Neurosci.* **13**, 119 (2019).
18. Rich, E. L. & Shapiro, M. L. Prelimbic/infralimbic inactivation impairs memory for multiple task switches, but not flexible selection of familiar tasks. *J. Neurosci.* **27**, 4747–4755 (2007).
19. Oualian, C. & Gisquet-Verrier, P. The differential involvement of the prelimbic and infralimbic cortices in response conflict affects behavioral flexibility in rats trained in a new automated strategy-switching task. *Learn. Mem.* **17**, 654–668 (2010).
20. Dutech, A., Coutureau, E. & Marchand, A. Reinforcement learning approaches to instrumental contingency degradation in rats. In *Conférence Française de Neurosciences Computationnelles - NeuroComp 2010* (eds Fourcaud-Trocmé, N. et al.) Vol. 105, 36–44 (Journal of Physiology-Paris, 2010).
21. Ottenheimer, D. J., Hjort, M. M., Bowen, A. J., Steinmetz, N. A. & Stuber, G. D. A stable, distributed code for cue value in mouse cortex during reward learning. *eLife* **12**, RP84604 (2023).
22. Otis, J. M. et al. Prefrontal cortex output circuits guide reward seeking through divergent cue encoding. *Nature* **543**, 103–107 (2017).
23. Schultz, W. Dopamine reward prediction error coding. *Dialogues Clin. Neurosci.* **18**, 23–32 (2016).
24. Hjort, M. Source data for 'Hjort et al. Prefrontal to Ventral Tegmental Area Dynamics Drive Contingency Degradation'. *figshare* https://doi.org/10.6084/m9.figshare.31431814 (2026).
25. Chen, T.-W. et al. Ultra-sensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300 (2013).
26. Hjort, M. et al. Microprisms enable enhanced throughput and resolution for longitudinal tracking of neuronal ensembles in deep brain structures. *Neurophotonics* **11**, 033407 (2024).

27. Spellman, T., Svei, M., Kaminsky, J., Manzano-Nieves, G. & Liston, C. Prefrontal deep projection neurons enable cognitive flexibility via persistent feedback monitoring. *Cell* **184**, 2750–2766 (2021).
28. Steinmetz, N. A., Zatka-Haas, P., Carandini, M. & Harris, K. D. Distributed coding of choice, action and engagement across the mouse brain. *Nature* **576**, 266–273 (2019).
29. Packer, A. M., Russell, L. E., Dalgleish, H. W. P. & Häusser, M. Simultaneous all-optical manipulation and recording of neural circuit activity with cellular resolution in vivo. *Nat. Methods* **12**, 140–146 (2015).
30. Marshel, J. H. et al. Cortical layer–specific critical dynamics triggering perception. *Science* **365**, eaaw5202 (2019).
31. Cohen, J. Y., Haesler, S., Vong, L., Lowell, B. B. & Uchida, N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* **482**, 85–88 (2012).
32. Gongwer, M. W. et al. Brain-wide projections and differential encoding of prefrontal neuronal classes underlying learned and innate threat avoidance. *J. Neurosci.* **43**, 5810–5830 (2023).
33. Hui, M. & Beier, K. T. Defining the interconnectivity of the medial prefrontal cortex and ventral midbrain. *Front. Mol. Neurosci.* **15**, 971349 (2022).
34. Zhuo, Y. et al. Improved green and red GRAB sensors for monitoring dopaminergic activity in vivo. *Nat. Methods***21**, 680–691 (2024).
35. van der Meulen, J. A. J., Joosten, R. N. J. M. A., de Bruin, J. P. C. & Feenstra, M. G. P. Dopamine and noradrenaline efflux in the medial prefrontal cortex during serial reversals and extinction of instrumental goal-directed behavior. *Cereb. Cortex* **17**, 1444–1453 (2007).
36. Gallello, I. et al. Dopamine responses in medial frontal cortex are more consistent with a generalized arousal signal than signed reward prediction errors. *J. Neurosci.* **46**, e1454252026 (2026).
37. Babiczky, Á & Matyas, F. Molecular characteristics and laminar distribution of prefrontal neurons projecting to the mesolimbic system. *eLife* **11**, e78813 (2022).
38. Watabe-Uchida, M., Eshel, N. & Uchida, N. Neural circuitry of reward prediction error. *Annu. Rev. Neurosci.* **40**, 373–394 (2017).
39. Tan, K. R. et al. GABA neurons of the VTA drive conditioned place aversion. *Neuron* **73**, 1173–1183 (2012).
40. van Zessen, R., Phillips, J. L., Budygin, E. A. & Stuber, G. D. Activation of VTA GABA neurons disrupts reward consumption. *Neuron* **73**, 1184–1194 (2012).
41. Preview : Journal Type Jeong, M. et al. Distinct interneuronal dynamics selectively gate target-specific cortical projections in drug seeking. *Neuron*https://doi.org/10.1016/j.neuron.2026.01.002 (2026).
42. Gallistel, C. R., Fairhurst, S. & Balsam, P. The learning curve: Implications of a quantitative analysis. *Proc. Natl Acad. Sci. USA* **101**, 13124 (2004).
43. Grossman, C. D., Bari, B. A. & Cohen, J. Y. Serotonin neurons modulate learning rate through uncertainty. *Curr. Biol.* **32**, 586–599 (2022).
44. Soltani, A. & Izquierdo, A. Adaptive learning under expected and unexpected uncertainty. *Nat. Rev. Neurosci.* **20**, 635–644 (2019).
45. Pearce, J. M. & Hall, G. A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol. Rev.* **87**, 532–552 (1980).
46. Mackintosh, N. J. A theory of attention: variations in the associability of stimuli with reinforcement. *Psychol. Rev.* **82**, 276–298 (1975).
47. Le Pelley, M. E. The role of associative history in models of associative learning: a selective review and a hybrid model. *Q. J. Exp. Psychol. B* **57B**, 193–243 (2004).
48. Coddington, L. T., Lindo, S. E. & Dudman, J. T. Mesolimbic dopamine adapts the rate of learning from action. *Nature* **614**, 294–302 (2023).
49. Behrens, T. E. J., Woolrich, M. W., Walton, M. E. & Rushworth, M. F. S. Learning the value of information in an uncertain world. *Nat. Neurosci.* **10**, 1214–1221 (2007).
50. Bissonette, G. B. & Roesch, M. R. Neural correlates of rules and conflict in medial prefrontal cortex during decision and feedback epochs. *Front. Behav. Neurosci.* **9**, 266 (2015).
51. Malagon-Vina, H., Ciocchi, S., Passecker, J., Dorffner, G. & Klausberger, T. Fluid network dynamics in the prefrontal cortex during multiple strategy switching. *Nat. Commun.* **9**, 309 (2018).
52. Karlsson, M. P., Tervo, D. G. R. & Karpova, A. Y. Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty. *Science* **338**, 135–139 (2012).
53. Powell, N. J. & Redish, A. D. Representational changes of latent strategies in rat medial prefrontal cortex precede changes in behaviour. *Nat. Commun.* **7**, 12830 (2016).
54. Rich, E. L. & Shapiro, M. Rat prefrontal cortical neurons selectively code strategy switches. *J. Neurosci.* **29**, 7208–7219 (2009).
55. Beier, K. T. et al. Circuit architecture of VTA dopamine neurons revealed by systematic input-output mapping. *Cell* **162**, 622–634 (2015).
56. Watabe-Uchida, M., Zhu, L., Ogawa, S. K., Vamanrao, A. & Uchida, N. Whole-brain mapping of direct inputs to midbrain dopamine neurons. *Neuron* **74**, 858–873 (2012).
57. Goldstein, R. Z. & Volkow, N. D. Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. *Nat. Rev. Neurosci.* **12**, 652–669 (2011).
58. Le, T. M., Potvin, S., Zhornitsky, S. & Li, C.-S. R. Distinct patterns of prefrontal cortical disengagement during inhibitory control in addiction: A meta-analysis based on population characteristics. *Neurosci. Biobehav. Rev.* **127**, 255–269 (2021).
59. Izquierdo, A. & Jentsch, J. D. Reversal learning as a measure of impulsive and compulsive behavior in addictions. *Psychopharmacology (Berl.)* **219**, 607–620 (2012).
60. Lüscher, C., Robbins, T. W. & Everitt, B. J. The transition to compulsion in addiction. *Nat. Rev. Neurosci.* **21**, 247–263 (2020).
61. Devoto, F., Zapparoli, L., Spinelli, G., Scotti, G. & Paulesu, E. How the harm of drugs and their availability affect brain reactions to drug cues: a meta-analysis of 64 neuroimaging activation studies. *Transl. Psychiatry* **10**, 429 (2020).

Methods

The University of Washington Institutional Animal Care and Use Committee approved all procedures in this manuscript under protocol no. 4450-01.

Subjects and surgery

All in vivo recordings occurred in male and female wild-type C57BL6J mice between 3 months and 6 months of age obtained from Jackson Laboratories or their progeny. Mice were housed in a 12-h reverse cycle, with experiments conducted during the dark period. The facility temperature was maintained between approximately 20 °C and 24 °C and at least 30% humidity. The task characterizations in Fig. 1 represent the behaviour of the first 20 mice investigated—a combination from Figs. 2 and 4. We added an additional 20 mice for model benchmark comparison including mice from Figs. 3 and 4 and controls in Fig. 5 and Extended Data Fig. 5. The results in Fig. 2 represent 15 mice, Fig. 3 contains 8 mice and the dopamine recordings in Fig. 4e represent 7 mice from an original cohort of 8 (1 was omitted because of fibre loss). The mPFC→VTA terminal recordings in Fig. 4q represent eight mice (NC), seven mice (saline) and seven mice (SCH + RAC), with one saline and one SCH + RAC omitted through outlier testing for reversal midpoint ($z = 2.07$ and 2.26, respectively, both $P < 0.05$). The mPFC→VTA single-cell recordings in Fig. 4f represent four mice. The ex vivo optogenetic results in Fig. 5a–g are from three VGAT-Cre (*VGAT* is also known as *Slc32a1*; Jax #028862) and three DAT-Cre (*DAT* is also known as *Slc6a3*; Jax #006660) mice. The in vivo optogenetics results in Fig. 5h–p derive from 20 mice, split randomly between ChR2+ (10) and mCherry (10) groups, and the inhibition results in Extended Data Fig. 5 are from 10 mice, split randomly into DREADD (5) and mCherry (5) groups. Sample sizes were not predetermined but are in line with other publications involving PFC single-cell imaging or neuronal photostimulation. For experiments with control and treatment groups, we split cage mates evenly into both groups during surgical preparation and used male and female mice for all groups in all studies. For experiments with control and treatment groups, blinding was not possible because the same experimenters performed the surgery (injecting different viruses or drug cocktails) and ran the behavioural analyses. Analysis of the behaviour of these cohorts was performed using established modelling programs that require minimal user input. Mice were housed in a reverse light-dark cycle and were group housed before surgery, after which they were housed singly. At least 1 week before the start of behaviour, mice experienced a water restriction of 1–2 ml per day[22,62] while maintaining at least 85% of their bodyweight. In addition to researchers, facility veterinarians monitored the animals' condition carefully.

Surgical implantation of microprisms centred on prelimbic cortex occurred following previously published protocols[26] with the following modifications and notes. Two-photon recordings in Fig. 2 occurred through 3-mm length prisms, whereas recordings for Figs. 3 and 4 occurred through 8-mm length prisms because of implant availability. The mPFC→VTA single-cell cohort (Fig. 4f–i) received bilateral injection of 500 nl AAVretro-pgk-cre (University of North Carolina vector core) into the VTA centred at anteroposterior −2.8, mediolateral ±0.13 and dorsoventral −4.0 at a 10° angle, and 4 × 300 nl of AAVDJ-DIO-GCaMP6s-WPRE (Stanford vector core) in mPFC following published coordinates[26]. The mPFC→VTA inhibition experiments (Extended Data Fig. 5) followed the same injection coordinates but received AAV5-DIO-CaMKIIa-hM4Di-mCherry-WPRE (University of Washington vector core) or AAV5-DIO-CaMKIIa-mCherry-WPRE (University of Washington vector core) in the mPFC. The mPFC→VTA fibre photometry recordings took place through 6-mm 400-µm core fibres (Doric MFC_400/470–0.37_6mm_MF2.5_FLT) implanted at the same coordinates, concurrently with bilateral injection of AAVDJ-GCaMP6s-WPRE (Stanford vector core) in mPFC at anteroposterior +1.94, mediolateral −0.4 and dorsoventral −2.3. Cannula inhibition experiments (Fig. 4k–q) included the same viruses and fibres, with the addition of a bilateral guide cannula (outer diameter 0.48 mm, length of cannula(C) = 3.5 mm, G1 = 0.5 mm G2 = 0.5 mm, RWD catalogue no. 62024) paired with dummy cannula (RWD catalogue no. 62124) and internal cannula (RWD catalogue no. 62224). Dopamine photometry recordings (Fig. 4a) involved the same fibres implanted at anteroposterior +1.94, mediolateral −0.4 and dorsoventral −2.4 after injection of AAV2/9-hSyn-GRAB-DA3h (BrainVTA)[34] at the same anteroposterior and mediolateral coordinates but dorsoventral −2.3. PFC-VTA stimulation mice (Fig. 5h) received a 300-nl bilateral injection of AAV5-CaMKIIa-hChR2(H134R)-eYFP-WPRE (University of Washington) or AAV5-CaMKIIa-mCherry-WPRE (University of Washington) at anteroposterior +1.94, mediolateral −0.4, dorsoventral −2.4, and bilateral implant of 200-µm core fibres (RWD R-FOC-BL200C-39NA) at 2.8, mediolateral ±1.3, dorsoventral −4.0 at a 10° angle. All viral titres were on the order of 1×10^{12} genome copies ml^{-1}, diluted in PBS if necessary. See Extended Data Fig. 3 for implant placements from animals in all experiments. Mice recovered from surgery for at least 3 weeks before they began water restriction, behavioural testing and neural recordings.

Head-fixed behaviour

Mice were head-fixed using a previously published apparatus[13,22,26] mounted to a goniometer stage (Thorlabs). They habituated to head-fixation and consumption of 2.5 µl 10% sucrose from a spout centred under their snout for at least three sessions before behaviour began. The sucrose concentration and volume were kept consistent across recordings. All behaviour hardware was custom built and Arduino Mega controlled[21], with animal performance saved though serial communication and custom Python scrips, similar to other systems[22,63]. Within the behavioural paradigm (Fig. 1a), an Aurora 206 A olfactometer delivered odours at 10% flow rate and 800SCCM overall flow rate of medical air[21]. The four odours diluted to 10% in mineral oil were selected randomly from a set of eight with neutral valence[64,65] and included butanol (banana), limonene (lime), carvone (mint), benzaldehyde (almond), geraniol (floral), ethyl butyrate (pineapple), 3-hexenal (grass) and pinene (pine).

Two cues began as high value (85% probability of 2.5 µl 10% sucrose reward), and two as low value (15% probability of 2.5 µl 10% sucrose reward). After mice displayed stable task performance, defined as anticipatory licking on at least 85% of 85% probability trials for 2 days, the contingencies of one high value (H→L) and one low value (L→H) cue were reversed, whereas the others remained constant (H→H, L→L). Cues were delivered from 1.5 s with a 1-s trace interval before reward delivery or omission. The inter-trial interval between reward delivery and the next cue was 16–20 s, sampled on a uniform distribution. Throughout this work, cues are denoted A→B, where A indicates the reward probability before the reversal, and B after. Mice trained to stable task performance following the new contingencies, after which they were perfused for placements following previous protocols[21] except in Fig. 3 in which mice experienced optogenetic manipulation before expiry.

We chose to focus our experiments on the first reversal because this allowed the best isolation of contingency degradation without the interaction of other processes that may arise from repeated toggling between learned contingency sets. There is considerable literature supporting the notion that the first reversal is not the same as subsequent, repeated reversal switches between learned contingency sets, and that mPFC is most involved in the initial contingency degradation before serial reversals between intertwined with set memory. Specifically, if animals are repeatedly overtrained on serial reversals and then mPFC is lesioned, studies repeat no impairment on the speed of subsequent reversals[18,66–69] or a muted effect that decreases over the number of serial reversals[19,54,70]. Only performing experiments during the first reversal allowed us to limit the contamination of contingency degradation signals in our recordings from those of other learned set memory processes.

Most of the behavioural analysis is described in 'Value modelling: RPE and meta-RPE'. Beyond the value modelling, 'normalized' licking represents the average number of pre-reversal anticipatory licks subtracted from the trial licks, and subsequently dividing between the difference of mean pre- and post-reversal anticipatory licks, smoothed with scipy.signal.gaussian(1). Scipy.optimize.curve_fit generated sigmoid (Fig. 1d,m) functions that described changes in licking behaviour, or relationships between licking and other variables (Fig. 1m). The sigmoid 'midpoint' (Figs. 1e, 4o and 5n and Extended Data Fig. 5e) is the inflection point of the resulting curve. The 'start' of sigmoid decay (Figs. 1m, 4o and 5n and Extended Data Fig. 5e) describes when the function decays more than 0.1 normalized licks or value units after reversal. Matplotlib generated most graphs, with seaborn utilized for heatmaps. GraphPad Prism v.10 calculated population statistics on bar graphs using either paired *t*-test, repeated-measures one-way analysis of variance (ANOVA) or two-way ANOVA, where appropriate, with Holm–Sidak post hoc correction for multiple comparisons (Supplementary Table 1). The colour schemes represented in this work utilize MetBrewer's Redon Pandora 1914 palette[71]. Our brain schematics were drawn from images originally in the Allen Brain Atlas at mouse.brain-map.org. See 'Single-cell imaging and optogenetics' and 'Fibre photometry, optogenetics, dopamine receptor pharmacology and DREADDs' for alignment of behavioural and neural data.

Value modelling: RPE and meta-RPE

Rescorla–Wagner (RPE) value models provided a canonical estimate of how an animal's internal valuation of stable and reversed cues changed throughout the task. Following a published protocol[72], an animal's 'choices' (presence or absence of anticipatory licking above baseline) were fit individually to each of the four reversal cues. The cue probability (0.85 or 0.15) provided value for time point 0. The RPE model returned an estimated value (Q_t, equation (1)) for each trial × cue type, in addition to how the outcome differed from the prediction (equation (3)), also known as the RPE. The comparison of these RPE estimates with behaviour is represented in Fig. 1f,h. The 'pre' model involved the first pre-reversal stable day, the 'rev' model the second pre-reversal stable day and reversal day(s) until the first post-reversal stable day. The 'post' model represents the second post stable day. For ease of comparison to the mRPE model, the RPE model[6,7] is reproduced below.

$$Q_{t+1} = Q_t + \alpha \times \delta_t \tag{1}$$

$$\delta_t = r_t - Q_t \tag{2}$$

The meta-RPE model is built off a *Q*-learning backbone, as it updates a value estimate for the next trial (Q_{t+1}, equation (3) using single-trial RPE (δ_t, equation (2)), static learning rate (α) and binary ({0,1}) reward outcome (r_t). The new flexibility related additions to the model come in the form of the overall meta-RPE on the trial (M_t, equation (4)) which itself models as a linear-rectifier balancing accumulated positive (CE, equation (5)) or negative (CD, equation (6)) errors, which rolling average at rate *e* and *d*, respectively. As in the RPE model, Q_0 initializes at the start reward probability for a cue and all error terms begin at 0. In total, the model, is thus:

$$Q_{t+1} = Q_t + \alpha \times \delta_t \times M_t \tag{3}$$

$$\delta_t = r_t - Q_t$$

$$M_t = \frac{-\mathrm{CD}_t + |\mathrm{CD}_t|}{2} + \frac{\mathrm{CE}_t + |\mathrm{CE}_t|}{2} \tag{4}$$

$$\mathrm{CD}_t = (1-d) \times \mathrm{CD}_{t-1} + d \times \delta_t \tag{5}$$

$$\mathrm{CE}_t = (1-e) \times \mathrm{CE}_{t-1} + e \times \delta_t \tag{6}$$

The BIC weighs goodness of fit against the number of free parameters[72], accounting for their influence on potential fit differences in the RPE model (two free parameters—the second free parameter comes from the stochasticity term in the softmax choice rule used to fit animal choice behaviour the value function[72]) and the mRPE model (four free parameters). Even with this penalty, the flexibility model outperforms the standard (Fig. 1l). The online rolling average of positive and negative RPEs at integration rates *d* and *e*, respectively (equations 5 and 6), allows the meta-RPE model to behave like the mice, with more stable choices until there is a concentration of repeated errors (Fig. 1i), at which time the value function begins to change (Fig. 1j). Separate CD (equation (5)) and CE (equation (6)) terms account for the respective differences in the rate of behavioural change towards repeated positive (CE) and negative (CD) RPEs (Fig. 1e). The addition of the linear rectification unit (equation (4)) balances these two types of error behaviour so that they influence only the value curve (equation (3)) when relevant, allowing CD to drive the change when the averaged error is negative, and CE to drive when the averaged error is positive. When the averaged error is close to 0, the value function is updated very little. That mPFC (Fig. 2c) meaningfully encodes only CD supports the decision to split the error accumulation in this way. Although the dopamine released into mPFC (Fig. 4f) displayed signatures of CE, the neural substrate for CE, and how VTA balances the influx of CD and CE, are open questions for future study.

We also compared the mRPE model with other models with dynamic learning rates including an RPE model with separate learning rates for positive (α_2) and negative errors (α_1) ($\mathrm{RPE}_{2\alpha}$, equation (7)),

$$Q_{t+1} = Q_t + \alpha_1 \times \delta_t^- + \alpha_2 \times \delta_t^+ \tag{7}$$

an RPE model that scales learning rate based on past trial type error history (RPE_{t-1}, equation (8)),

$$Q_{t+1} = Q_t + \alpha \times \delta_t \times \delta_{t-1} \tag{8}$$

and salience models including PH and eMack[45–47]. The PH and eMack models also split impacts of positive and negative errors to produce a dynamic learning rate. Notably, the eMack model increases learning rate as predictors become more reliable in their prediction of an outcome (as occurs during CE), whereas the PH model increases learning rate as predictors become more uncertain (as occurs during CD). Both models also use two competing value functions for action–outcome and action–no outcome that need not sum to 1. See Le Pelley[47] for additional description alongside the full PH and eMack equations. See Extended Data Fig. 1b for model curves in an example animal.

That the mRPE model outperforms all other models we considered indicates that its performance derives from the shared impact of several characteristics (Extended Data Fig. 1c): a dynamic alpha (shared with PH, eMack and RPE_{t-1}) that splits (shared with $\mathrm{RPE}_{2\alpha}$, PH and eMack) and averages (shared with PH) positive and negative errors over time to increment a single value function (shared with RPE, RPE_{t-1}, $\mathrm{RPE}_{2\alpha}$). As models containing some but not all of these characteristics do not outperform the mRPE model, the characteristics are all probably important to produce the best fit to CD behaviour and cannot individually explain the superior performance of the mRPE model.

Single-cell imaging and optogenetics

All two-photon calcium imaging took place through microprisms in mPFC visualizing GCaMP6s (see 'Subjects and surgery' for more details) at 7.5 Hz and 920 nm excitation. Owing to availability, data sets were collected on different microscopes, but the microscope was held consistent within an experiment. The dataset in Fig. 2 was collected on an Olympus FVMPE-RS[22] using an Olympus XLPLN ×10 immersion objective and a MaiTai DeepSee tunable Ti:Sapphire laser. Figure 3 data were collected on a Bruker 2p+ with SLM module and

Cousa objective[73], using a tunable IR laser (Insight, Spectra-Physics) and a fixed wavelength 1,040 nm photostimulation laser (Spirit, Spectra-Physics). A Bruker Investigator with a Coherent Chameleon laser and Cousa objective collected the mPFC→VTA imaging in Fig. 4. Visualizing imaging planes at 820 nm, GCaMP6's approximate isosbestic wavelength[74], assisted manual plane registration across days[26]. For the optogenetic stimulation experiments in Fig. 3, target masks were selected manually after plane matching by comparing the live image with a target averaged image from the reversal day with GLM masks overlaid for the excited CD cells (SLM experiments 1–3), GLM^- cells (SLM experiment 2), or GLM^+ cue cells (SLM experiment 3). Masks were an 8-μm diameter spiral stimulated for 10 s at 20 Hz with a 37.5 ms duty cycle (Fig. 3e), following previously reported power levels[26].

After collection, Suite2p motion corrected all non-optogenetic recordings within and across days[26,75]. Owing to difficulties with the shutter closed artefact, Suite2p did not correct successfully the stimulation data in Fig. 3, but TurboReg[76] in ImageJ proved an adequate substitution. If across day tracking was unsuccessful, somatic ROIs were curated manually and tracked in ImageJ. Before more advanced analysis, a second-order Butterworth filter with cut-off at 0.1 Nyquist frequency lightly denoised data, which was then z-scored to ease comparison across data sets. This maintained peak integrity, as is visible in Fig. 2b. Custom scripts aligned the extracted ROIs to behaviour for further analysis. Cells then entered the GLM pipeline ('Generalized linear modelling') for sorting. Peri-event time histograms (Fig. 2g) for cell visualization were normalized to a 2.5 s baseline before cue onset. For activity classification (Figs. 2g and 4h), excited refers to a mean z-scored activity between 0 s and 8 s post cue larger than zero, and inhibited refers to mean activity less than or equal to zero. There was no significant difference (Extended Data Fig. 4j) in the population difference of excited and inhibited cells split by animal when comparing Figs. 2g and 4h. Overlap maps within (Fig. 2e) or across (Extended Data Fig. 4) days represent binary mask comparison (numpy.intersect1d) for thresholded GLM variables.

Generalized linear modelling

GLMs are a statistical tool useful for parsing the relationships between predictors and a neural signal. Of particular concern to this study was the tight temporal relationship between relevant cognitive variables and sensorimotor responses. For example, during the cue period the sensory property of the cue odour itself, the value prediction from the cue–reward association, and the anticipatory licking motor response all co-occur. Given the dominance of motor signals across cortical areas[77], careful separation of the contribution of each signal is necessary to isolate cognitive variables, such as value or mRPE.

Within a given session, we model the ith neuron as

$$Y = X\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{9}$$

where Y is GCaMP6s activity for T time points for an individual neuron, X is a $(T \times p)$ matrix composed of submatrices $X^{(j)}$ for $j = 1, 2, \ldots, J$, which are $(T \times m_j)$ matrices such that $\sum_j m_j = p$. The p-vector $\boldsymbol{\beta}$ comprise the coefficients in the linear model and $\boldsymbol{\epsilon}$ is a length T vector of error terms. We are interested in testing a null hypothesis of the form $H_{0j}: \beta_j = 0$, under which the model of interest becomes the reduced model

$$Y = X_{-j}\beta_{-j} + \epsilon, \tag{10}$$

where X_{-j} and β_{-j} indicate the omission of columns and entries corresponding to the variables according to $X^{(j)}$. After fitting the model using ordinary least squares, the observed F-statistic is defined by

$$F = \frac{(\mathrm{RSS}_R - \mathrm{RSS}_F)/m_j}{\mathrm{RSS}_F/(T-p)}, \tag{11}$$

where RSS_R is the residual sum of squares under the reduced model and RSS_F is the residual sum of squares under the full model.

Standard GLM packages assume independence of the observations. When this assumption is violated (as in the temporal dependence setting of our paper), the F-statistic given above is meaningful, in the sense that it will tend to take on a small value when $H_{0j}: \beta_j = 0$ holds and a larger value when the null hypothesis is violated. However, the P values arising from standard GLM packages will vastly overstate the statistical evidence against H_{0j} (that is, the P values will be much too small) in the presence of temporal dependence. This is because the F-statistic does not follow the 'theoretical' distribution $F_{m_j,(T-p)}$ under H_{0j}.

This motivated us to empirically generate a null distribution that considers the temporal dependency of the data. Our approach is outlined in 'Algorithm 1'. Specifically, in step 1, we repeatedly time-wrap (circularly shift[78]) Y to preserve the temporal structure of the data while destroying association between Y and X, as the shifted data is misaligned from the task parameters that could predict it (Extended Data Figs. 2 and 3). For the bth time-wrapped dataset, for $b = 1, \ldots, B$, we record F^{*b}, the corresponding F-statistic for $H_{0j}: \beta_j = 0$. Then, in step 2, a P value associated with $H_{0j}: \beta_j = 0$ is computed by comparing the F-statistic from a model properly aligned to the task predictors (F) to the empirical distribution of the F-statistic arising from the repeatedly circularly shifted dataset (F^{*b}). The P value is the fraction of circularly shifted data sets for which the empirical F-statistics from circularly shifting, $F^{*1}, \ldots, F^{*B}$, exceed the observed F-statistic F (Extended Data Fig. 2).

We note that in step 1, to obtain a sufficiently rich empirical null distribution $F^{*1}, \ldots, F^{*B}$, we circularly shift[78] the full set of neurons recorded during this session, rather than only the ith neuron. Repeatedly circularly shifting only a single neuron does not lead to a rich enough null distribution of the F-statistic, as repeated circular shifts of only a single observed neuron trace can be somewhat repetitive when B is large. Incorporating all observed neurons in the circular shifting results in a distribution $F^{*1}, \ldots, F^{*B}$ that is more representative of the full spectrum of calcium dynamics.

For fibre photometry data, we took a similar approach, where we again fit a linear model, but now Y is either the summed neural activity, or the dopamine signal, for all of the neurons under the fibre. However, to generate a null distribution, we circularly shift data from the fluorescent indicator from all of the recording days and for all subjects. This amounts to a small modification of 'Algorithm 1', where i in step 1 indexes a random day and a random subject, rather than a random neuron from that session.

Finally, after setting a significance threshold ($P < 0.01$ for neural data, or $P < 0.05$ for photometry data owing to lower signal to noise), we either rejected or failed to reject $H_{0j}: \beta_j = 0$. This procedure was repeated for each neuron in the case of single-cell two-photon imaging, and each fibre in the case of fibre photometry, and for each task predictor(s) (that is, each β_j, for $j = 1, \ldots, K$).

Algorithm 1. Compute a P value for significance of chunk $X^{(j)}$ for neuron i

Step 0: define the 'circular shift' operation

Define CircularShift(Y, T, s):

$$Y^* \leftarrow [Y[(s+1:T), Y[1:s]]]$$

return(Y^*)

Step 1: sample the null distribution of the F-statistic using circular shifting

for b in $\{1, 2, \ldots, B\}$ do

$i_b^* \leftarrow$ random index for a neuron recorded in the session

$s \leftarrow$ random time index from 1 to T
$Y_b^* \leftarrow$ CircularShift($Y_{i_b^*}$, T, s)
$\mathrm{RSS}_F \leftarrow$ Residual sum of squares in full model using Y_b^* and X
$\mathrm{RSS}_R \leftarrow$ Residual sum of squares in reduced model using Y_b^* and X

$$F^{*b} \leftarrow \frac{(\mathrm{RSS}_R - \mathrm{RSS}_F)/m_j}{\mathrm{RSS}_F/(T-p)}$$

end for
Step 2: compare the observed F-statistic against the null distribution
$\mathrm{RSS}_F \leftarrow$ residual sum of squares in full model using Y_i and X
$\mathrm{RSS}_R \leftarrow$ residual sum of squares in reduced model using Y_i and X

$$F \leftarrow \frac{(\mathrm{RSS}_R - \mathrm{RSS}_F)/m_j}{\mathrm{RSS}_F/(T-p)}$$

$P[(\#\text{ of } F^{*b} > F) + 1]/[B + 1]$.

The following sets of variables comprised $X^{(j)}$ for $j = 1, 2, \ldots, K$ used to predict the neural signal Y: cue (H→L, H→H, L→H, L→L on stable days, or a single combined kernel on reversal days as representations could be unstable as contingencies shifted), licks (shifted forward and backward in time 300 ms to account for preparatory feedback-related encoding), sucrose delivery (1 s) or omission (1 s), value (equation (3)), RPE (equation (2)), CE (equation (6)) and CD (equation (5)). The data in this manuscript speak to the utility of this method to meaningfully characterize neural signals.

Fibre photometry, optogenetics, dopamine receptor pharmacology and DREADDs

Fibre photometry recordings occurred on a TDT RZ10 following methods described previously[63]. In brief, A 465 nm LED excited both indicators, and a 405 nm LED provided isosbestic correction for GCaMP6s recordings (Fig. 4p,q), but as the isosbestic wavelength of GRAB-DA3h was not well characterized at the time and headfixation minimizes motion artefacts[63], a correction was not performed for the dopamine dataset (Fig. 4a–e). After collection, each signal underwent photobleaching correction through a fifth degree polynomial and denoising using the OASIS algorithm[79], adapted to fit published indicator kinetics[34] where necessary. Peri-cue time histogram construction occurred comparably with that described in 'Single-cell imaging and optogenetics' and GLM fitting took place as described in 'Generalized linear modelling'. The Sal and SCH + RAC cohorts (Fig. 4k–q) received cannula infusions 15 min before recording at a concentration of 30 ng SCH and 300 ng RAC[80] in 5% dimethylsulfoxide in saline per site per day.

Optogenetic stimulation of the mPFC→VTA projection took place bilaterally with a 470-nm laser (SLOC Model BL473T8-150) at 20 Hz, 10 ms pulse width and 5 mW power, checked daily. For the real-time place preference experiment (Fig. 5b), animals habituated to a two-chamber arena[63] for 10 min before recording began. EthoVision then quantified the animal's body centroid (Fig. 5b) for 15 min, with stimulation triggered automatically in the assigned zone. Behavioural hardware for the task stimulation experiment differed from the standard head-fixed setup ('Head-fixed behaviour') only in the addition of a transistor–transistor logic line that triggered the onset of stimulation for H→L trials. The mRPE model fits in Fig. 5h represent individual models trained for each day, with Q_0 at 0.85.

For chemogenetic inhibition experiments we used the hM4Di designer receptor activated exclusively by designer drugs (DREADD[81,82]). All mice that were part of the DREADD inhibition experiment in Extended Data Fig. 5 received intraperitoneal saline 15 min before the task throughout acquisition, and across reversal in the saline group. The DREADD group received 1 mg kg^{-1} injection of deschloroclozapine[83] (in 1% dimethylsulfoxide in saline) on reversal days −1 to 3. After this, DREADD mice received saline injections until their post-reversal behaviour stabilized and the experiment ended. The more muted effect from this experiment is consistent with predictions of the mRPE model (where animals can still learn from alpha and RPE without mRPE) and mPFC lesion literature where lesions impair but do not abolish reversal[18,19,66,70,84–86].

Ex vivo two-photon calcium imaging of VTA

VGAT-Cre and DAT-Cre mice underwent stereotaxic surgery under isoflurane anaesthesia (3–4% induction, 0.75–1.5% maintenance). To enable optogenetic stimulation of mPFC inputs and Cre-dependent calcium imaging in VTA GABAergic (VGAT-Cre) and dopaminergic (DAT-Cre) neurons, all mice received injections of AAV8-hsyn-ChRmine-Kv2.1 into the mPFC (anteroposterior +1.94, mediolateral ±0.5, dorsoventral −2.2) and AAV1-hsyn-FLEX-GCaMP6f into the VTA (angle 10°, anteroposterior −2.8, mediolateral ±1.22, dorsoventral −4.4).

For ex vivo recordings and imaging, mice were anaesthetized deeply with Euthasol (0.06 ml 30 g^{-1}, intraperitoneally) and decapitated. Brains were removed rapidly and immersed for 2 min in ice-cold, carbogenated *N*-methyl-D-glucamine (NMDG)-based artificial cerebrospinal fluid (ACSF; 92 mM NMDG, 2.5 mM KCl, 1.25 mM NaH_2PO_4, 30 mM $NaHCO_3$, 20 mM HEPES, 25 mM glucose, 2 mM thiourea, 5 mM sodium ascorbate, 3 mM sodium pyruvate, 0.5 mM $CaCl_2{\cdot}4H_2O$, 10 mM $MgSO_4{\cdot}7H_2O$, 12 mM *N*-acetyl-L-cysteine; pH 7.3–7.4). Coronal midbrain slices (300 µm) were prepared on a vibratome (VT1200, Leica), incubated for 15 min at 34 °C in NMDG ACSF and then transferred to HEPES-based ACSF (same composition but containing 2 mM $CaCl_2{\cdot}4H_2O$ and 2 mM $MgSO_4{\cdot}7H_2O$) at room temperature for 30 min before imaging.

Optogenetic stimulation of ChRmine-expressing mPFC terminals in the VTA was delivered using a pE-100 LED system (CoolLED; 615 nm). Stimulation trains were generated using a Master-9 pulse generator (A.M.P.I.) and consisted of 20 pulses (20-ms pulse duration, 20 Hz), preceded by a 30-s baseline period, at 2.90 mW power output.

Two-photon calcium imaging was performed using an Olympus FluoView FVMPE-RS microscope equipped with a scientific CMOS camera (QImaging) at an acquisition rate of 1 Hz. Imaging data were processed and analysed using ImageJ, Suite2p and custom Python scripts.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data are available at GitHub (https://github.com/stuberlab/Hjort-et-al.-2026-PFC-and-reversal-learning) and at Figshare (https://doi.org/10.6084/m9.figshare.31431814)[24].

Code availability

Code is available at GitHub (https://github.com/stuberlab/Hjort-et-al.-2026-PFC-and-reversal-learning) and at Figshare (https://doi.org/10.6084/m9.figshare.31431814)[24].

62. Guo, Z. V. et al. Procedures for behavioral experiments in head-fixed mice. *PLoS ONE* **9**, e88678 (2014).
63. Gordon-Fennell, A. et al. An open-source platform for head-fixed operant and consummatory behavior. *eLife* **12**, e86183 (2023).
64. Lee, J., Linster, C. & Devore, S. Odor preferences shape discrimination learning in rats. *Behav. Neurosci.* **127**, 498–504 (2013).
65. Saraiva, L. R. et al. Combinatorial effects of odorants on mouse behavior. *Proc. Natl Acad. Sci. USA* **113**, E3300–E3306 (2016).
66. De Bruin, J. P. C. et al. Role of the prefrontal cortex of the rat in learning and decision making: effects of transient inactivation. *Prog. Brain Res.* **126**, 103–113 (2000).
67. Seamans, J. K., Floresco, S. B. & Phillips, A. G. Functional differences between the prelimbic and anterior cingulate regions of the rat prefrontal cortex. *Behav. Neurosci.* **109**, 1063–1073 (1995).
68. Hervig, M. E. et al. Dissociable and paradoxical roles of rat medial and lateral orbitofrontal cortex in visual serial reversal learning. *Cereb. Cortex* **30**, 1016–1029 (2020).

69. Dalton, G. L., Wang, N. Y., Phillips, A. G. & Floresco, S. B. Multifaceted contributions by different regions of the orbitofrontal and medial prefrontal cortex to probabilistic reversal learning. *J. Neurosci.* **36**, 1996–2006 (2016).
70. Boulougouris, V., Dalley, J. W. & Robbins, T. W. Effects of orbitofrontal, infralimbic and prelimbic cortical lesions on serial spatial reversal learning in the rat. *Behav. Brain Res.* **179**, 219–228 (2007).
71. Mills, B. GitHub - BlakeRMills/MetBrewer: color palette package in R inspired by works at the Metropolitan Museum of Art in New York. *GitHub* https://github.com/BlakeRMills/MetBrewer?tab=readme-ov-file (2023).
72. Rhoads, S. A. & Gan, L. Computational models of human social behavior and neuroscience: An open educational course and Jupyter Book to advance computational training. *J. Open Source Educ.* **5**, 146 (2022).
73. Yu, C.-H. The Cousa objective: a long-working distance air objective for multiphoton imaging in vivo. *Nat. Methods* **21**, 132–141 (2024).
74. Barnett, L. M., Hughes, T. E. & Drobizhev, M. Deciphering the molecular mechanism responsible for GCaMP6m's $Ca2^+$-dependent change in fluorescence. *PLoS ONE* **12**, e0170934 (2017).
75. Pachitariu, M. et al. Suite2p: beyond 10,000 neurons with standard two-photon microscopy. Preprint at *bioRxiv* https://doi.org/10.1101/061507 (2017).
76. Thevenaz, P., Ruttimann, U. E. & Unser, M. A pyramid approach to subpixel registration based on intensity. *IEEE Trans. Image Process.* **7**, 27–41 (1998).
77. Engel, T. A. & Steinmetz, N. A. New perspectives on dimensionality and variability from large-scale cortical dynamics. *Curr. Opin. Neurobiol.* **58**, 181–190 (2019).
78. Harris, K. D. Nonsense correlations in neuroscience. Preprint at *bioRxiv* https://doi.org/10.1101/2020.11.29.402719 (2021).
79. Friedrich, J., Zhou, P. & Paninski, L. Fast online deconvolution of calcium imaging data. *PLoS Comput. Biol.* **13**, e1005423 (2017).
80. de Jong, J. W. et al. A neural circuit mechanism for encoding aversive stimuli in the mesolimbic dopamine system. *Neuron* **101**, 133–151 (2019).
81. Roth, B. L. DREADDs for neuroscientists. *Neuron* **89**, 683–694 (2016).
82. Armbruster, B. N., Li, X., Pausch, M. H., Herlitze, S. & Roth, B. L. Evolving the lock to fit the key to create a family of G protein-coupled receptors potently activated by an inert ligand. *Proc. Natl Acad. Sci. USA* **104**, 5163–5168 (2007).
83. Nagai, Y. et al. Deschloroclozapine, a potent and selective chemogenetic actuator enables rapid neuronal and behavioral modulations in mice and monkeys. *Nat. Neurosci.* **23**, 1157–1167 (2020).
84. Kosaki, Y. & Watanabe, S. Dissociable roles of the medial prefrontal cortex, the anterior cingulate cortex, and the hippocampus in behavioural flexibility revealed by serial reversal of three-choice discrimination in rats. *Behav. Brain Res.* **227**, 81–90 (2012).
85. Brigman, J. L. & Rothblat, L. A. Stimulus specific deficit on visual reversal learning after lesions of medial prefrontal cortex in the mouse. *Behav. Brain Res.* **187**, 405–410 (2008).
86. Chudasama, Y. & Robbins, T. W. Dissociable contributions of the orbitofrontal and infralimbic cortex to Pavlovian autoshaping and discrimination reversal learning: further evidence for the functional heterogeneity of the rodent frontal cortex. *J. Neurosci.* **23**, 8771–8780 (2003).
87. Paxinos, G. & Franklin, K. B. J. *Paxinos and Franklin's the Mouse Brain in Stereotaxic Coordinates* (Academic Press, 2019).

Acknowledgements We thank Y. Li, K. Deisseroth, C. Ramakrishnan and L. Zweifel for viral constructs. We thank B. Briones for suggesting the two-way ANOVA visualization, R. Chang for assistance with photometry pilot experiments, A. Campuzano for histology assistance and C. Zhou for maintaining the two-photon microscopes. We also thank P. Phillips and S. Golden for their feedback and advice. This work was supported by F31DA053706 and T32EB031512 (M.M.H.); R25DA057786 (Z.G.); K99DA059612 (A.G.G.); U01NS113252, the Pew Biomedical Scholars Program, and the Klingenstein-Simons Fellowship in Neuroscience (N.A.S.); the Murdock Foundation and P30DA048736 (G.D.S. and M.R.B.); R37DA033396 (M.R.B.) as well as NSF193428 and R37DA032750 (G.D.S.).

Author contributions M.M.H. and G.D.S. are responsible for the conceptualization of the project and designed the study. M.M.H. and N.A.S. developed the mRPE model. E.A. and G.D.S. wrote the GLM software, and M.M.H. and G.D.S. wrote the learning model comparison notebook. Many authors were involved in the investigation phase alongside M.M.H.: G.D.S. contributed to Fig. 5b–g; Z.Q.G. to Figs. 4b–e,m–q, 5h–o and Extended Data Fig. 3; A.G.G. to Figs. 4m–q, 5b–o and Extended Data Fig. 5a–e; P.Y.L. to Fig. 5h–o and Extended Data Figs. 3 and 5a–e; and M.T. to Fig. 5b–i and Extended Data Fig. 3. M.M.H. performed curation, visualization and formal analysis. G.D.S., N.A.S. and M.R.B. provided resources. M.M.H. wrote the original draft, which she reviewed and edited alongside G.D.S. with input from all authors. G.D.S. supervised the project. M.M.H., G.D.S. and M.R.B. acquired funding for this work.

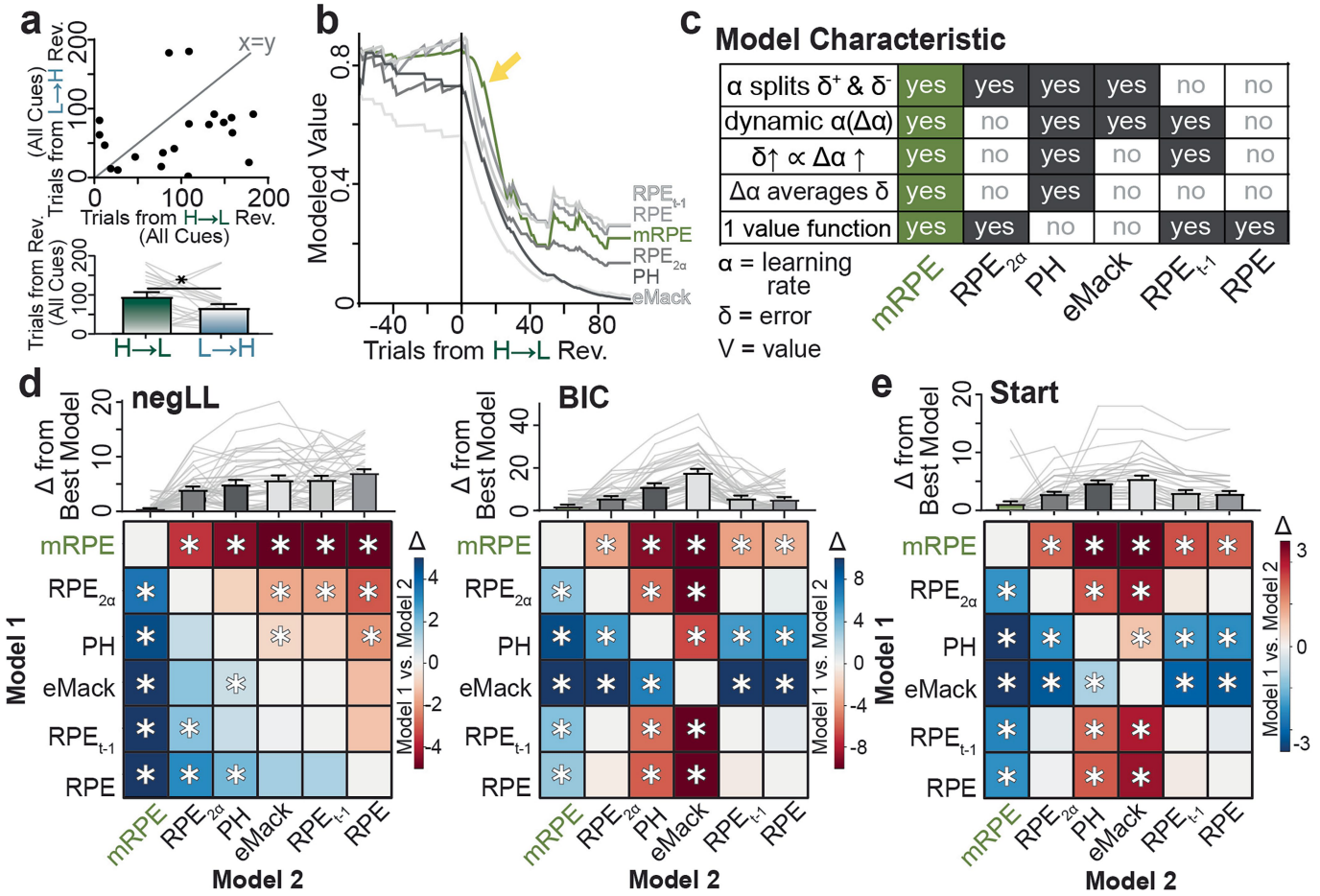
Competing interests The authors declare no competing interests.

Additional information

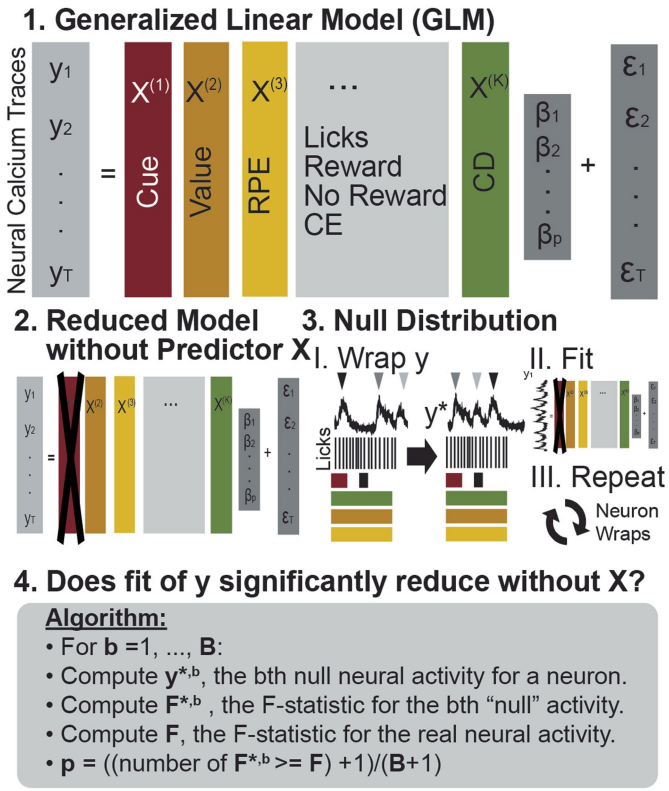
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-026-10443-5.

Correspondence and requests for materials should be addressed to Garret D. Stuber.

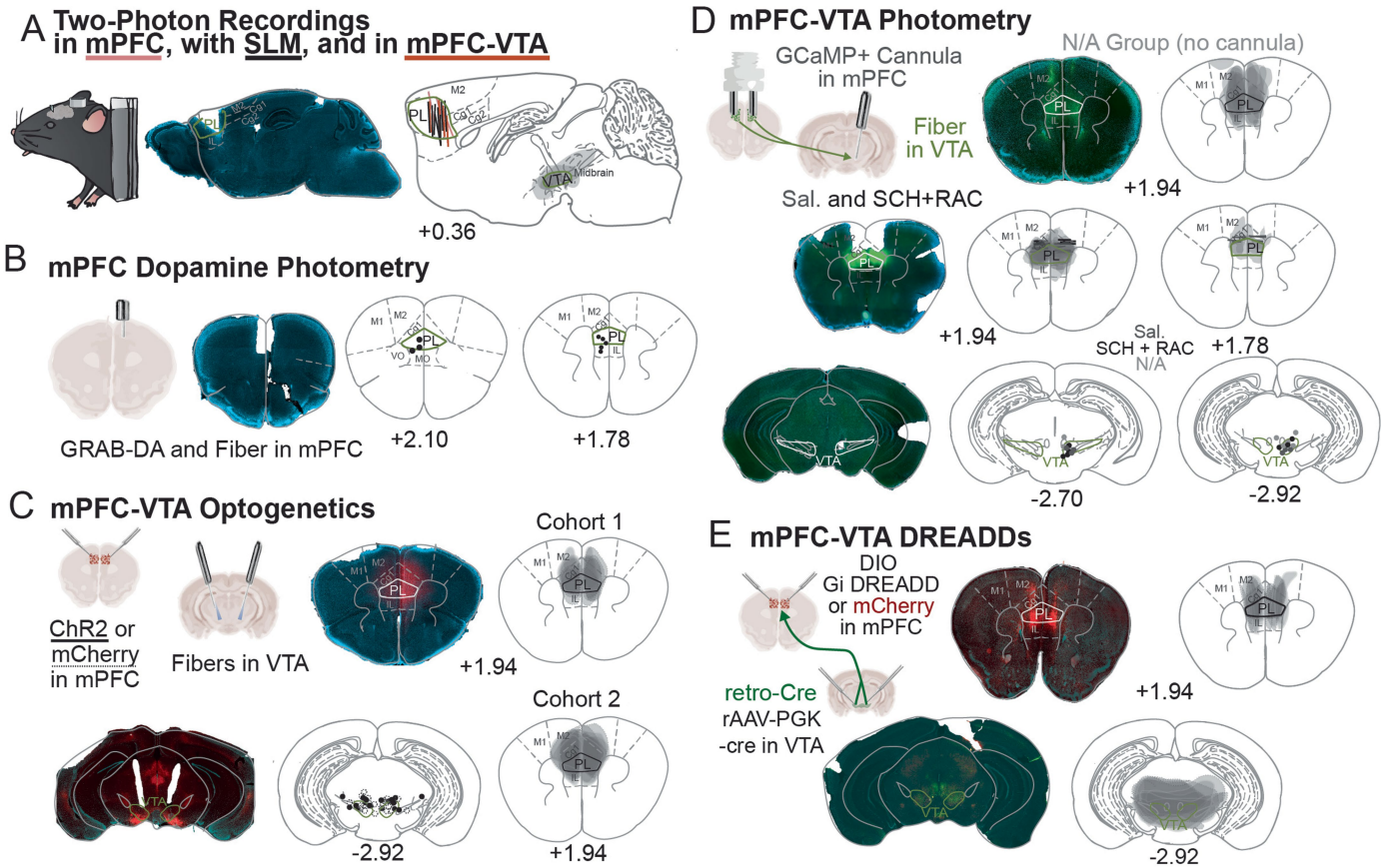
Peer review information *Nature* thanks Armin Lak and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



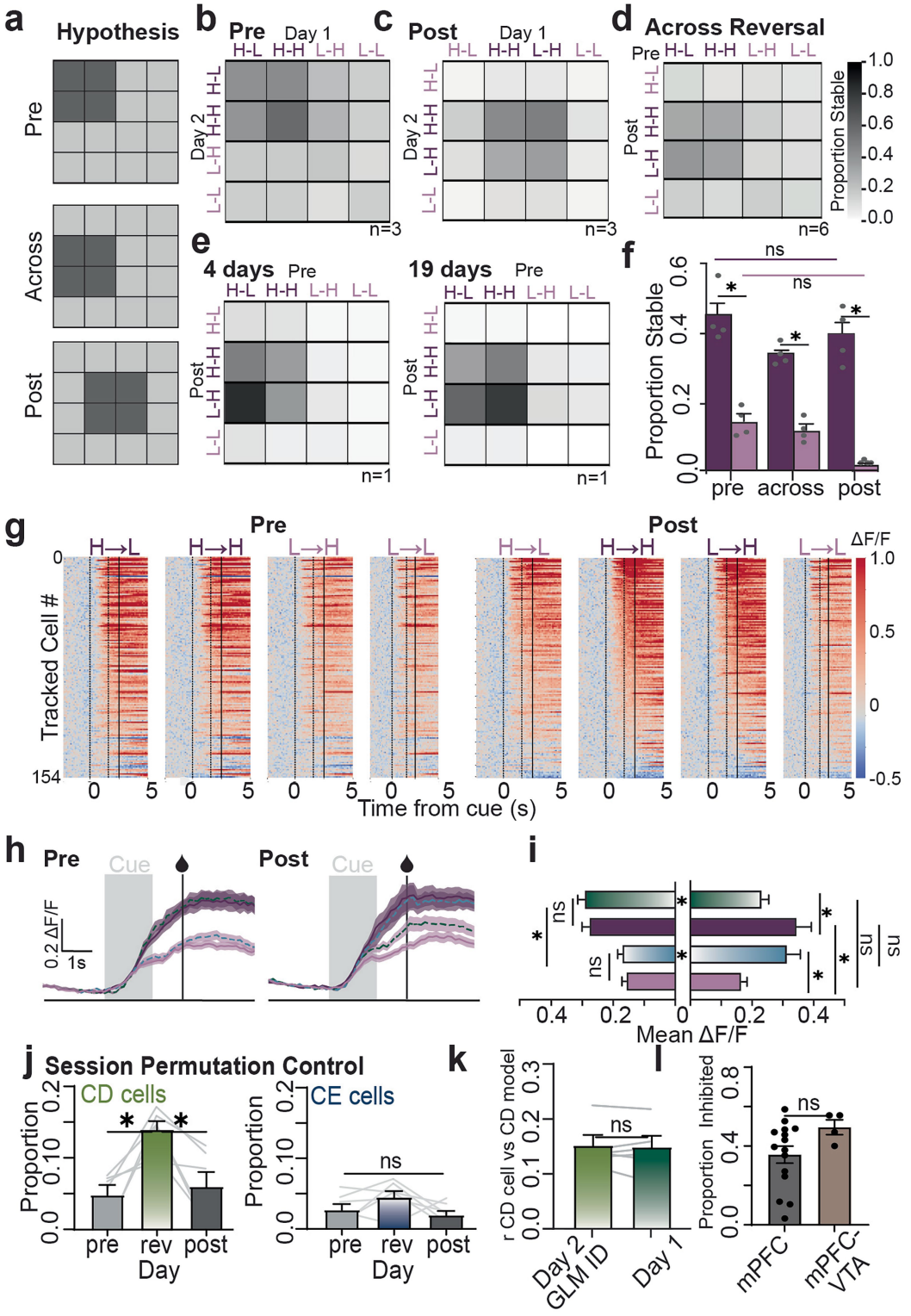
Extended Data Fig. 1 | Additional mRPE Model Quantifications. a. like Fig. 1e but in the context of all trials for all cues, instead of just the reversed cue. The H→L reversal still takes significantly longer than the L→H reversal on average ($p = 0.0352$, $n = 20$ animals). **b**. Visualization of different model fits for an example animal, note how the mRPE model function (green) takes longer to begin to decay after the reversal (yellow arrow). **c**. Table comparing model characteristics between the models in Fig. 1k–m. **d**. Unweighted (negLL) and weighted (BIC) fit comparisons between models, quantified as the change in score from the best of the 6 models (top), with post-hoc comparisons visualized (lower). The heat plots quantify the mean difference between model 1 (right vertical) and model 2 (left lower). Lower scores (redder) indicate better model fit. Asterisks indicate significant post-hoc differences at $p < 0.05$ ($n = 40$ animals). This panel supports Fig. 1k & l. **e**. Like ***d***, but comparing the start of reversal (decay at least 10% from pre-reversal level) between models. See Fig. 1m for direct comparison to behavior. All error bars: mean + /− SEM. Asterisk (*) indicates statistical significance at $p < 0.05$, see Supp. Table 1 for more statistical information, including more post-hoc comparisons, sidedness, and corrections for multiple comparisons.



Extended Data Fig. 2 | Schematic of Time-Series Generalized Linear Modelling. This figure depicts the process of assigning significant encoding of specific predictors in neurons. Please see Methods- 'Generalized linear modelling' for more details.

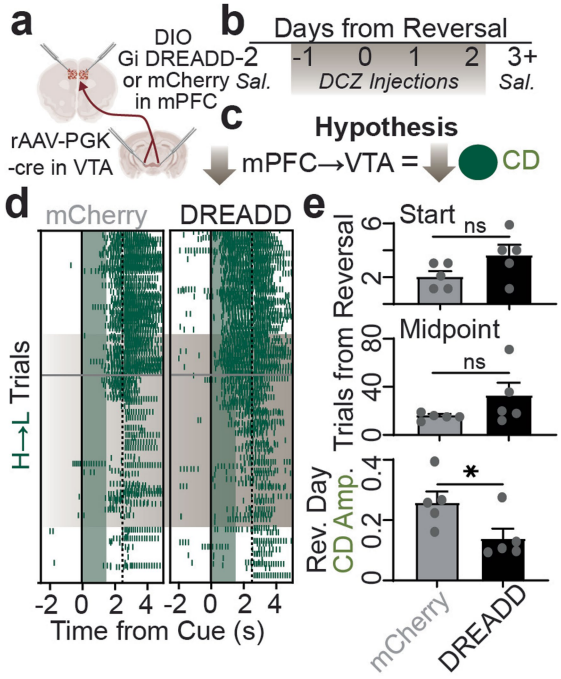


Extended Data Fig. 3 | Implant Placement Histology. a. Schematic of two-photon preparation (left), with sample image (center), and placements for all experiments (right). Placements from experiments in Fig. 2 only are in pink (n = 3/6), Fig. 3 in black (n = 8/8), and Fig. 4 in orange (n = 4/4). The spread of the retro-cre virus (as in Fig. 4l) is denoted with gray translucent overlay for applicable mice. The tissue from three mPFC (pink) placements was inadvertently destroyed in an unsuccessful histology pilot and therefore not included. **b**. mPFC dopamine photometry schematic as in Fig. 4a (left) with sample placement (center) and implant map (left) from all mice (n = 7/7). **c**. Optogenetic recording schematic, as in Fig. 5a, with virus injection in mPFC and bilateral fiber placement over VTA (top left), with sample viral spread image (top center), and combined spread map from all ChR2+ (solid border, n = 10/10) and mCherry (dashed border, n = 10/10) animals (top right). Also included are sample VTA fiber tip placements (lower left) and hit maps from ChR2+ (n = 9/10, solid black) and mCherry (n = 10/10, dashed outline) animals (lower right). One ChR2+ animal's VTA slices were too damaged to localize targeting and are omitted. **d**. mPFC→VTA terminal recording strategy with GCaMP6s injected in mPFC and photometry fiber placement over VTA, as in Fig. 4k (top left) with sample mPFC GCaMP6s spread image (top center) and spread maps from all N/A animals (n = 8/8) with translucent overlay (top right). VTA placements (n = 8/8 n/a, 7/7 Sal. 7/7 SCH + RAC) are in the lower row, with a sample fiber tip track (lower left) and hit maps (lower right). Bilateral cannula locations (lines) for Sal. and SCH + RAC mice alongside GCaMP6s spread are also provided (center). **e**. Schematic of mPFC-VTA DREADD expression strategy, as in Extended Data Fig. 5a (top left) alongside sample DREADD expression in mPFC (top center) and hit maps (top right) (n = 5/5 DREADD, 5/5 mCherry). Also included are retro-Cre spread maps for all mice (lower). Abbreviations: PL= prelimbic cortex, IL = infralimbic cortex, Cg1 = dorsal anterior cingulate cortex (ACC), Cg2 = ventral ACC. PL, IL, and ACC comprise the rodent medial prefrontal cortex[15]. M1 = primary motor cortex, M2 = secondary motor cortex, MO =medial orbital cortex, LO = lateral orbital cortex, VTA = ventral tegmental area, as defined in the Paxinos Atlas[87].



Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Stability across Contingency Reversal. a. Hypothesis for cue stability cell proportions, from previous work[21] compared to neural results before (pre), after (post), and across reversal. **b**. Cue stability map on two days before reversal, where H → L and H → H cues are high value (dark purple). **c**. Cue stability map constructed from two post-reversal stable behavior days, where H → H and L → H cues are high value (dark purple). **d**. Across reversal cue stability map, with cells tracked from pre-reversal (x-axis) where H → L and H → H are high value to post-reversal (y-axis) where H → H and L → H cues are high value. **e**. Example animals with different spacing between stable pre and post day compared across reversal. **f**. Statistics on the proportion of stable cue cells for high value (dark purple) and low value (light purple) cues before, after, and across reversal. There is not a significant difference in the proportion of coding within a cue type in any of the conditions ($p = 0.1715$, data from $n = 4$ cross-cue comparisons from $n = 6$ animals). **g**. Trial-averaged time histogram of stable cue cell activity before (pre) and after (post) contingency reversal, sorted by H → H activity. **h**. Mean +/-SEM traces for stable cue cells on H → L (green dash), H → H (dark purple solid), L → H (light blue dash) and L → L (light purple solid). Error bars are shaded in accordance with high (dark purple) or low (light purple) value. **i**. Quantification of average stable cue cell activity during cue and trace interval (0–2.5 s) split by cue (Cue x time $p = 0.0032$, cells from $n = 6$ animals). There is not a significant difference within coding of 85% or 15% cues before or after reversal. **j**. Session permutation control results, which reflect the correlates of modeled mRPE signals from reversal days evaluated with GLM on neurons from reversal and other days. If CD neurons displayed a similar activity profile to the reversal days on pre- or post- days, there would not be a significant increase in the proportion of isolated cells on the reversal day (CD: $p = 0.0071$, CE: $p = 0.1923$, $n = 20$ animals). **k**. Reversal cell stability quantification. There is not a significant difference ($p = 0.6970$ $n = 6$ animals) in the relationship with the CD signal for tracked cells on the identification day (2) vs the previous day (1). **l**. There is not a significant difference between the proportion of excited/inhibited cells in imaging in Fig. 2g vs Fig. 4h ($p = 0.1287$, $n = 4$ PFC-VTA, $n = 15$ PFC animals). See Supp. Table 1 for more statistical information, including more post-hoc comparisons, sidedness, and corrections for multiple comparisons. See Extended Data Fig. 3 for placements.



Extended Data Fig. 5 | Inhibiting mPFC → VTA Signaling Impairs CD. **a**. Schematic of DREADD expression viral strategy **b**. Animals received deschloroclozapine, DCZ (1 mg/kg) one day before and three days during reversal. **c**. This experiment tests the hypothesis that inhibiting the mPFC→VTA projection will reduce CD signaling during reversal. **d**. Lick raster plots from example animals suggest a modestly impaired H → L reversal in Gi-DREADD animals (right) compared to mCherry controls (left). **e**. Quantifying this relationship reveals a significantly reduced CD amplitude from the mRPE model on the first reversal day ($p = 0.0238$, $n = 5$ animals per group). All error bars: mean +/− SEM. Asterisk (*) indicates statistical significance at $p < 0.05$, see Extended Data Table 1 for more statistical information, including more post-hoc comparisons, sidedness, and corrections for multiple comparisons. See Extended Data Fig. 3 for virus expression.

nature portfolio

Corresponding author(s): Garret Stuber

Last updated by author(s): Feb 27, 2026

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We used the following established programs for our data collection: Bruker Prairie View and Olympus Fluoview (2p data collection), Aurora 206A (olfactometer), TDT RZ10 (photometry), ImageJ/FIJI (2.17.0). We used custom Arduino/Python scripts (Python 3.7) to collect behavior data and synchronize it with recordings. These programs were previously reviewed and published and are available at https://github.com/agordonfennell/OHRBETS
Data analysis	We used the following established programs for our analysis: Suite2p (motion correction and ROI extraction), GraphPad Prism 10 (All statistics besides GLM). We used custom Python 3.7 scripts to analyze our data. Of particular relevance to this project is the python distribution of the p-value GLM package which is available at GitHub https://github.com/stuberlab/Hjort-et-al.-2026-PFC-and-reversal-learning. We also consider a number of behavioral analysis models, including our own in this work. A script that runs and compares all models is also available on GitHub alongside the full dataset of behavior from 40 animals

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The data generated in this paper are publicly available in a FigShare repository https://doi.org/10.6084/m9.figshare.31431814.
We aligned slices to the Paxinos Atlas, digitally available at ISBN 9780128161586

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	n/a
Reporting on race, ethnicity, or other socially relevant groupings	n/a
Population characteristics	n/a
Recruitment	n/a
Ethics oversight	n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were not predetermined but are in line with other publications involving PFC single cell imaging or neuronal photostimulation including Otis et al. 2017 (PMID: 28225752), Jennings et al. 2019 (PMID: 30651638), Ottenheimer et al. 2023 (PMID: 37382590), and Piantadosi et al. 2023 (PMID: 38086375). The breakdowns for each experiment are available on Supplementary Data Table 1.
Data exclusions	One animal was excluded from the dopamine fiber photometry experiment (Fig. 4 a-e) due to fiber loss during data collection, and one additional animal was included in pre/post analysis (Fig. 4d) but excluded from the photometry GLM analysis (Fig. 4e) due to a detached patch cable during the reversal recording. One DREADD animal was excluded from ED Fig 5 due to headcap loss mid-experiment. Anticipating that a small fraction of animals may struggle with the reversal task, we performed outlier testing using the Grubbs Outlier test in GraphPrism for all animal behavior datasets (independently within each experimental condition) on the H-L and L-H reversal midpoints. Under this criteria, one control ($Z=2.03$, $p<0.05$) animal was excluded from the experiment in ED Fig. 5 due to significant outlier reversal midpoints
Replication	The major results of this manuscript were duplicated in at least two independent cohorts of multiple animals. For the model data in Fig. 1, we developed our model using n=20 mice and replicated our findings in n=20 additional mice. The imaging data in Fig. 2 comes from 4 independent cohorts of mice (total n=15 mice, split n=3, n=3, n=3, n=6) with consistent results. The SLM results in Fig 3 for Exp. 1 and 2 were replicated in two independent cohorts of animals (total n=8 animals, split n=3, and n=5). SLM Exp. 3 took place in only the second cohort, with consistent results across n=5 mice. Mice for the dopamine photometry experiment in Fig. 4a were split into 2 cohorts (total n=7, split n=3, n=4). The mPFC-VTA imaging experiment (Fig. 4f) was split into 2 cohorts (total n=4, split n=1, n=3). The mPFC-VTA photometry experiment was split into 2 no cannula (N/A) cohorts (total n=8, split n=4, n=4) and 2 experimental cohorts with cannula (total n=7 SCH+RAC, n=7 saline animals, split: cohort 1= 4 SCH-RAC, 3 Sal, cohort 2= 3 SCH-RAC, 4-saline). The Slice 2p experiments (Fig. 5a) represent data from n=15 GABA slices from n=3 vgat-cre animals and n=22 DA slices from n=3 dat-cre animals. The mPFC-VTA optogenetics experiments from Fig. 5h represent data from 2 cohorts (total n=10 ChR2 and n=10 mCherry animals, split: cohort 1=4 ChR2 and 4 mCherry, cohort 2=6 ChR2 and 6 mCherry). The post-reversal laser-off timepoint was only collected in cohort 2 (n=6 ChR2 and n=6 mCherry animals). The mPFC-VTA DREADD experiment (Fig. ED5) represents results from 2 cohorts (total n=5 DREADD and n=5 mCherry animals, split: cohort 1= 1 DREADD and 3 mCherry animals, cohort 2= 4 DREADD and 2 mCherry). In all cases, results replicated between animals within a cohort as well as across cohorts for each experiment.

Randomization	For experiments with control and treatment groups, we split cage mates evenly into both groups during surgical preparation and used male and female mice for all groups in all studies
Blinding	For experiments with control and treatment groups, blinding was not possible because the same experimenters performed the surgery (injecting different viruses or drug cocktails) and ran the behavior. Analysis of the behavior of these cohorts was performed using established modeling programs that require minimal user input.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Anti-Cre Recombinase Antibody, clone 2D8 MAB3120
Validation	This antibody has been validated by Sigma-Aldrich, the manufacturer as well as in over 150 peer-reviewed publications. More information is available on the maufacturer's site: https://www.sigmaaldrich.com/US/en/product/mm/mab3120

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	This study used male and female Wild-Type C57BL6J mice between 3 and 6 months of age obtained from Jackson Labs or their progeny, except for the slice 2p experiments in Fig. 5 which used the progeny of Vgat-Cre and DAT-Cre mice between 2 and 4 months of age.
Wild animals	No wild animals were used in this study.
Reporting on sex	We used a combination of male and female mice for all studies. We did not observe any sex differences in coding or behavioral outcomes.
Field-collected samples	This study did not include field collected samples.
Ethics oversight	The University of Washington Institutional Animal Care and Use Committee approved all procedures in this manuscript under protocol #4450-01

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks	n/a
Novel plant genotypes	n/a
Authentication	n/a