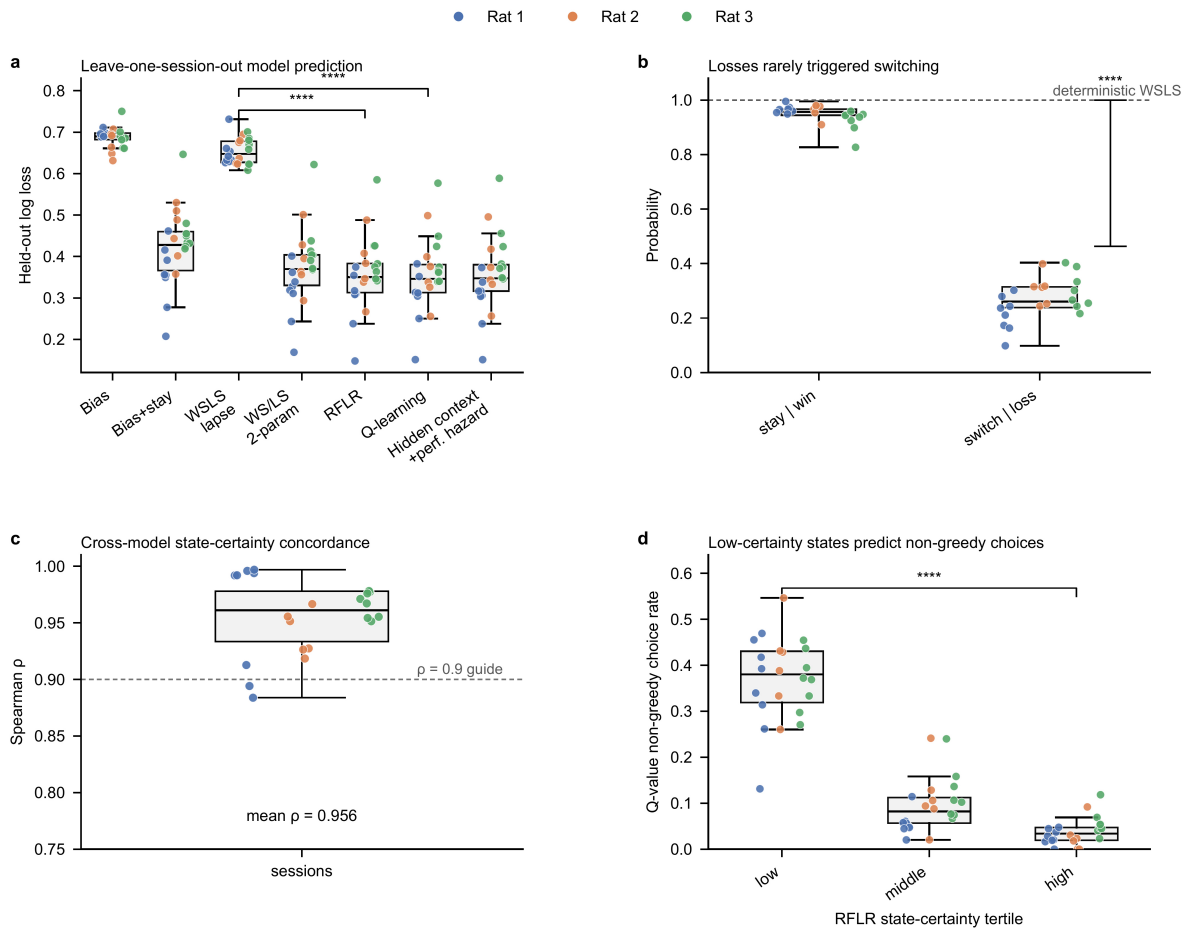
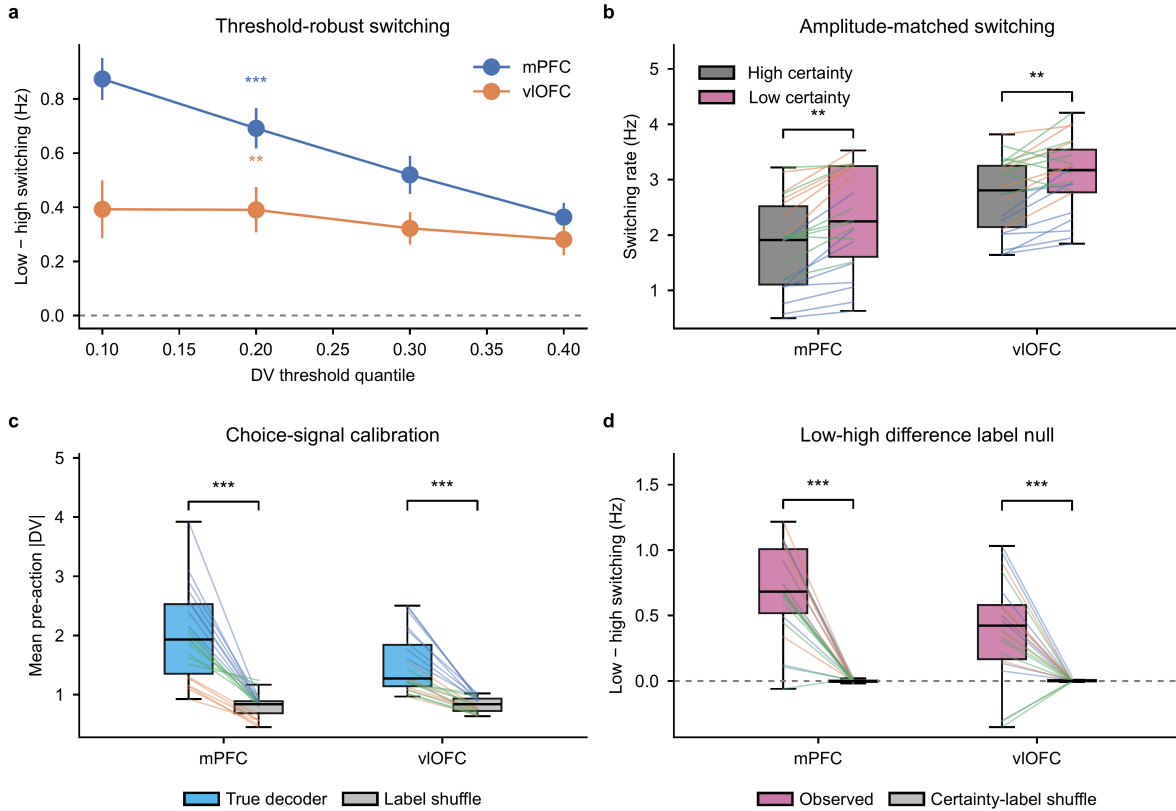


1 Supplementary



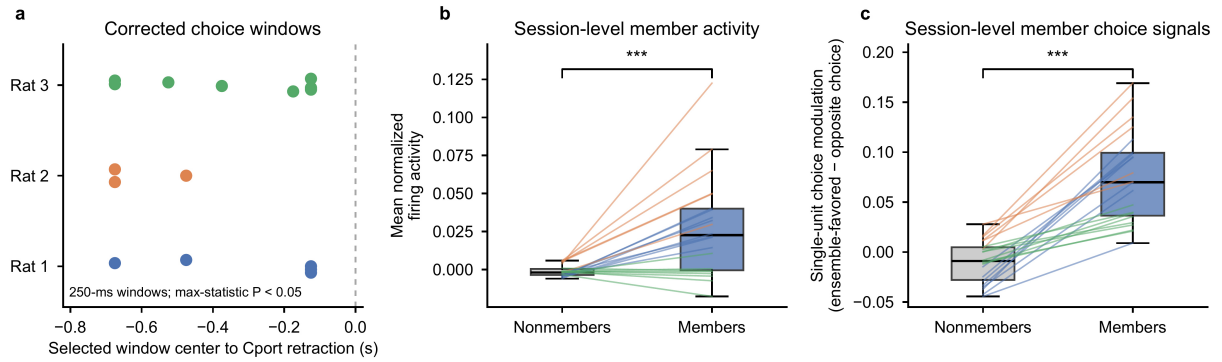
Supplementary Figure 1 | Behavioral model comparison supports RFLR-derived choice-state certainty.

a, Leave-one-session-out prediction performance for seven behavioral models fit separately for each rat. The hidden-context model inferred left-better versus right-better contexts using a Bayesian filter with a switch hazard that depended on a decaying accumulator of previous losses. Lower held-out log loss indicates better prediction. RFLR and Q-learning models outperformed a win-stay/lose-switch (WSLS) lapse model, whereas their held-out performance did not differ significantly. b, Descriptive WSLS behavior across sessions. Rats showed high stay-after-win probability but low switch-after-loss probability relative to the deterministic WSLS expectation of 1, indicating that unrewarded trials did not automatically trigger switching. c, Session-level concordance between RFLR and Q-learning state-certainty estimates, computed as the Spearman correlation between trial-wise RFLR certainty, $2|P_{\text{RFLR}}(\text{left}) - 0.5|$, and Q-learning certainty, $2|P_Q(\text{left}) - 0.5|$. The dashed line marks $\rho = 0.9$; mean $\rho = 0.956$ across sessions. d, Q-value non-greedy choice rate across within-session tertiles of RFLR state certainty. Non-greedy choices were defined as choices opposite to the action with the higher Q value from the held-out Q-learning model. Low RFLR-certainty states were enriched for Q-value non-greedy choices relative to high-certainty states. In all panels, boxes show session-level distributions and dots represent individual sessions colored by rat identity ($n = 22$ sessions, 3 rats). Significance was assessed using linear mixed-effects models with rat identity as a random effect; planned comparisons were Benjamini-Hochberg adjusted within panel (****, $q < 0.0001$).



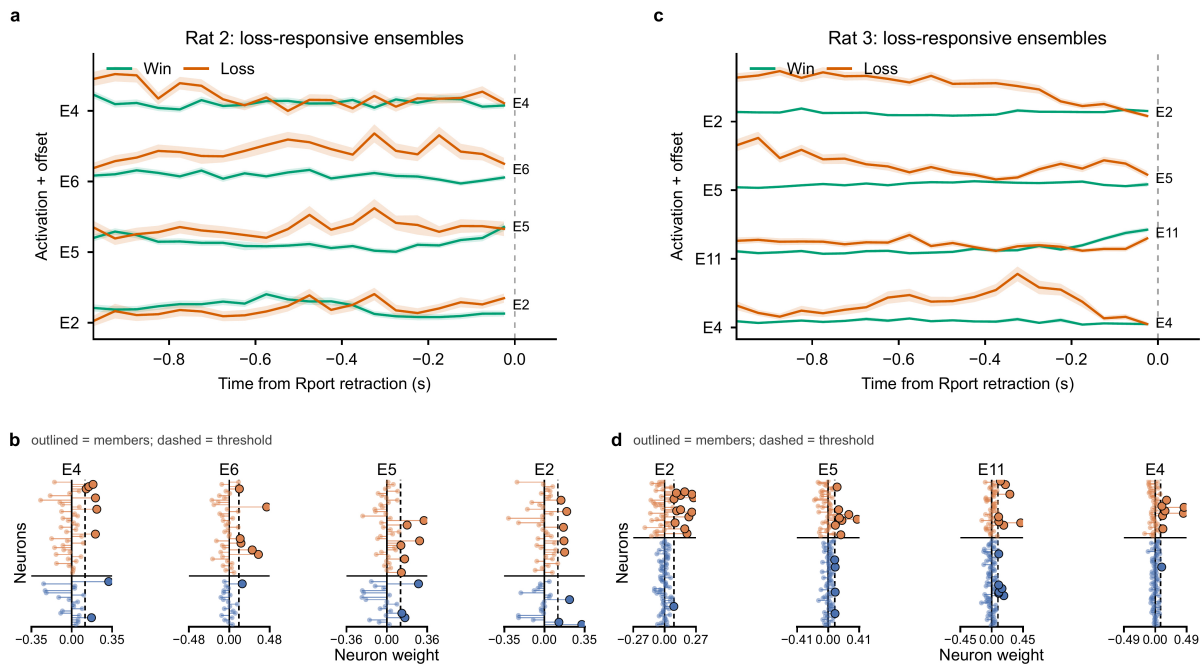
Supplementary Fig. 2 | Control analyses supporting rapid alternation of decoded choice representations in low-certainty states.

a, Switching rate was quantified from the posterior decision variable (DV; logit posterior for decoded upcoming choice) as transitions between stable positive and negative DV states during the pre-action window. DV thresholds were set from the 10th–40th percentiles of the empirical |DV| distribution pooled across trial-by-time bins within each session and region. Low-minus-high certainty differences remained positive across thresholds. **b**, High- and low-certainty trials were matched by mean pre-action |DV| amplitude before recomputing the switching rate. Low-certainty trials still showed higher switching rates in both mPFC and vIOFC, arguing against a simple weak-signal explanation. **c**, Mean pre-action |DV| from the true decoder was compared with decoders retrained after shuffling choice labels 500 times per session. Larger |DV| in the true decoder indicates that the DV captured genuine choice-related structure. **d**, The observed low-minus-high switching-rate difference was compared with a certainty-label-shuffle null, in which high/low certainty labels were shuffled within session while preserving decoded DV traces and group sizes. The observed difference exceeded this null in both regions, indicating that enhanced switching depended on the behavioral certainty-state assignment. Box plots summarize sessions; lines connect paired session values and are colored by rat identity. Error bars in **a** show mean $\pm$ SEM across sessions. Significance was assessed with linear mixed-effects models with rat identity as a random effect. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$; n.s., not significant.



Supplementary Fig. 3 | Validation of Cport-hold choice- and certainty-related ensembles.

a, Time windows in which inter-areal ensembles showed significant upcoming-choice selectivity during the central-port (Cport) hold period. Each dot denotes one selected ensemble; colours indicate rat identity. Choice selectivity was screened with 250-ms sliding windows and a within-session choice-label shuffle with max-statistic correction, retaining ensembles at $P < 0.05$. **b**, Mean normalized firing activity of high-weight ensemble members versus nonmembers, averaged by session across selected ensembles and their choice-coding windows. **c**, Single-unit choice modulation for high-weight ensemble members versus nonmembers. Modulation is defined as activity for the ensemble-favoured choice minus activity for the opposite choice, measured in the selected choice-coding window and averaged by session. In **b** and **c**, box plots summarize $n = 22$ paired sessions from three rats; lines connect paired session means and are coloured by rat identity. Significance was assessed with linear mixed-effects models with rat identity as a random effect. High-weight members showed higher normalized firing activity than nonmembers (**b**, $P = 2.2 \times 10^{-7}$) and stronger choice modulation (**c**, $P = 1.5 \times 10^{-23}$), supporting the weight-defined membership used to interpret Cport-hold ensembles.



Supplementary Fig. 4 | Loss-responsive vIOFC-dominant inter-areal ensembles in additional animals.

a,c, Activation traces of loss-responsive inter-areal ensembles from Rat 2 and Rat 3, aligned to reward-port retraction. Traces show mean $\pm$ s.e.m. across trials for rewarded (win) and unrewarded (loss) outcomes. Ensembles are vertically offset and ordered by the timing of loss-related activation. **b,d**, Neuron-weight distributions for the same ensembles shown in a and c, with panel order matching the activation traces. Blue and orange points denote mPFC and vIOFC neurons, respectively; the horizontal line separates regions. Outlined points indicate ensemble members, defined as neurons with weights exceeding the within-ensemble mean + 1 s.d.; dashed vertical lines mark this membership threshold. These additional examples show that loss-responsive inter-areal ensembles are present beyond the representative animal shown in the main figure and are dominated by high-weight vIOFC constituents.