

Medial Prefrontal and Orbitofrontal neural-ensemble interactions reflect decision updates differently upon surprises versus uncertainty

Mehmet Yanik

yanik@ethz.ch

Swiss Federal Institute of Technology in Zurich (ETH Zurich) <https://orcid.org/0000-0002-8963-2893>

Zhehan Fu

Swiss Federal Institute of Technology in Zurich (ETH Zurich)

Bifeng Wu

Harvard University

Tansel Yasar

Swiss Federal Institute of Technology in Zurich (ETH Zurich) <https://orcid.org/0000-0003-4424-0172>

Peter Gombkoto

Swiss Federal Institute of Technology in Zurich (ETH Zurich) <https://orcid.org/0000-0001-8393-173X>

Eminhan Özil

Swiss Federal Institute of Technology in Zurich (ETH Zurich)

Wolfgang von der Behrens

Swiss Federal Institute of Technology in Zurich (ETH Zurich) <https://orcid.org/0000-0001-6883-0236>

Article

Keywords:

Posted Date: October 16th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11011042/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Medial Prefrontal and Orbitofrontal neural-ensemble interactions reflect decision updates differently upon surprises versus uncertainty

Zhehan Fu^{1,2,*}, Bifeng Wu^{1,2,3,*}, Tansel Baran Yasar¹, Peter Gombkoto^{1,2}, Eminhan Özil^{1,2}, Wolfger von der Behrens^{1,2} & Mehmet Fatih Yanik^{1,2}

¹ Neurotechnology Group, Institute of Neuroinformatics, Department of Information Technology and Electrical Engineering, ETH Zurich and University of Zurich, Zurich, Switzerland

² Neuroscience Center Zurich, University of Zurich and ETH Zurich, Zurich, Switzerland

³ Present address: Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA

* These authors contributed equally: Zhehan Fu, Bifeng Wu.

Correspondence and requests for materials should be addressed to Mehmet Fatih Yanik (yanik@ethz.ch).

Abstract

Temporal integration and correct interpretation of anomalies are important for decision stability and flexibility. Their impairment is a hallmark of many cognitive disorders and has been linked to dysfunction of medial prefrontal cortex (mPFC) and orbitofrontal cortex (OFC). How these regions coordinate event-by-event preference updating following outcomes remains unclear. Using ultraflexible electrodes, we simultaneously tracked the same neurons in mPFC and ventrolateral OFC (vOFC) across days while rats performed a two-armed bandit task with unsignaled reward reversals. Neural choice representations were temporally stable during high-certainty states (inferred from exploitative behavior), but exhibited rapid beta-range-timescale alternations between competing choice signals during low-certainty states characterized by subsequent exploratory switching. Before action execution, choice-selective, certainty-modulated inter-areal ensembles were preferentially dominated by mPFC neurons and more strongly represented the eventual choice under high certainty. Upon losses, however, vOFC exhibited sequential activation of purely loss-specific ensembles throughout the post-outcome period. Furthermore, when losses followed high-confidence choices, vOFC-preceding mPFC outcome dynamics were selectively associated with subsequent switching. Thus, OFC ensembles monitor failures under both exploratory and exploitative conditions. Yet, pre-decision-confidence dependent coordination of post-decision mPFC–vOFC population helps interpret whether negative outcomes should be tolerated as stochastic feedback or treated as policy-relevant failures.

34 Adaptive behavior in dynamic and uncertain environments requires organisms to integrate
35 recent actions and outcomes in order to infer whether a negative outcome reflects stochastic
36 feedback or a meaningful change in action-outcome contingencies.^{1,2} Resolving this ambiguity
37 is essential for balancing exploitation of an established action policy against exploration of
38 alternatives. Policies should persist while current beliefs remain reliable but be revised when
39 accumulated evidence indicates environmental change. Impaired cognitive flexibility is
40 relevant across several psychiatric conditions,³ while obsessive–compulsive disorder provides
41 a specific example linked to altered frontostriatal connectivity and goal-directed planning.⁴
42 Human lesion studies, rodent electrophysiology and cross-species evidence implicate medial
43 prefrontal and orbitofrontal regions in reversal learning, action selection and flexible decision-
44 making.^{5–8} These findings motivate the hypothesis that interactions between mPFC and OFC
45 help determine when established action policies are maintained or updated. OFC has also
46 been proposed to represent latent task states, allowing observations to be interpreted
47 according to context.⁸

48 Determining how medial prefrontal cortex (mPFC) and orbitofrontal cortex (OFC) jointly
49 support adaptive decision-making requires distinguishing distributed representation from
50 interareal coordination. Behaviorally relevant variables can be represented across widely
51 distributed neuronal populations.⁹ Medial frontal populations can also be recruited flexibly
52 according to task demands. During memory-based choices, this recruitment is accompanied
53 by phase-locking between medial frontal neurons and amygdala and hippocampal field
54 potentials.¹⁰ Low-dimensional communication subspaces and direction-specific population
55 activity patterns provide candidate mechanisms for selective interactions between cortical
56 areas.^{11,12} Oscillatory coherence offers a complementary mechanism for regulating interareal
57 communication.^{13,14} Under this framework, decision stability and behavioral updating may
58 depend on coordinated dynamics across distributed networks. Simultaneous population
59 recordings can resolve cross-area coactivation and relative timing that recordings confined to
60 one region cannot provide.¹⁵ However, these measurements reveal statistical coordination and
61 temporal precedence; they do not independently establish causal or anatomically direct
62 information transfer.

63 Linking neural activity to latent decision states poses a methodological challenge because
64 these states are not directly observable and may be expressed across distributed neuronal
65 populations. Behavioral models provide one approach for inferring evolving action preferences
66 from recent choices and outcomes.¹⁶ Here, we use model-derived choice-state certainty as an
67 operational index of the relative strength of this inferred action preference, without equating it
68 with subjective confidence. Previous studies have linked OFC activity to decision confidence
69 and confidence-guided behavior,^{17,18} while medial OFC supports outcome retrieval when
70 relevant information must be inferred.¹⁹ However, these studies do not explain how latent
71 decision states are organized across interacting mPFC and OFC populations. Addressing this
72 gap requires methods that capture both concurrent interareal population structure^{11,12} and the
73 relative timing of regional activity patterns.^{15,20} Cross-correlation of trial-wise population-
74 decoding signals provides an analytical approach for estimating such timing differences
75 between regions.²¹ We therefore combined behavioral modeling with simultaneous mPFC–
76 vOFC recordings to test whether choice-state certainty organizes distributed ensemble
77 recruitment and interareal temporal dynamics.

78 To test this, we simultaneously recorded populations of single neurons from mPFC and
79 ventrolateral OFC (vOFC) while rats performed a probabilistic reversal learning task with

80 unsignaled contingency changes.^{22–24} We then asked how neural representations and inter-
81 areal interactions differed across the model-derived low- and high-certainty states. Model-
82 derived choice-state certainty was associated with distinct modes of distributed population
83 dynamics: high-certainty states exhibited stronger and more stable representations of
84 upcoming choice, whereas low-certainty states were characterized by rapid alternations
85 between competing choice representations. At the ensemble level, mPFC-dominated inter-
86 areal ensembles preferentially represented upcoming choices during high-certainty states,
87 whereas vOFC recruited loss-specific ensembles following unrewarded outcomes. After
88 losses occurring following high-confidence choices (i.e. surprises), vOFC outcome
89 representations preceded those in mPFC on trials followed by a choice switch. These findings
90 suggest that certainty-dependent interactions between distributed prefrontal populations
91 contribute to the differential maintenance and updating of behavioral policies under uncertainty
92 versus surprises.

93 Results

94 To examine how prefrontal populations support flexible decision-making under uncertainty,
95 we combined behavioral modeling with simultaneous single-unit recordings from medial
96 prefrontal cortex (mPFC) and ventrolateral orbitofrontal cortex (vOFC) while rats performed a
97 two-armed bandit reversal task.

98 **Rats flexibly update action preferences after unsignaled reversals and** 99 **enter model-graded choice states**

100 Rats performed a two-armed bandit reversal task using three nose ports: one central port
101 (Cport) for trial initiation and two lateral reward ports (Rports) for reporting choices and
102 collecting reward (25% sucrose solution, 50 μ L; Fig. 1a). Each trial began with illumination of
103 the Cport. Rats held their nose in the Cport until the LED turned off, then chose between the
104 two Rports. Within each block, one Rport was assigned a high reward probability (80%) and
105 the other a low reward probability (20%). This assignment remained fixed for 40–50 trials
106 before reversing without explicit cues. Because reward delivery was probabilistic, individual
107 outcomes were ambiguous: unrewarded high-probability choices and rewarded low-probability
108 choices both occurred. Sessions contained 250–400 trials (328 ± 13 , mean $\pm$ s.e.m.; 22
109 sessions, 3 rats).

110 Rats adapted reliably to these unsignaled contingency reversals. In a representative session,
111 choice preference gradually tracked the currently high-reward port, and this evolving
112 preference was captured by a recursively formulated logistic regression (RFLR) model (Fig.
113 1b). In this model, current choice probability reflects both choice persistence and a latent
114 evidence variable that integrates past choice–outcome interactions with exponential decay.
115 Across sessions, the probability of choosing the higher-reward port shifted toward the newly
116 advantageous side over approximately 10–15 trials after reversal (Fig. 1c). This gradual
117 adaptation, together with a low probability of switching after unrewarded trials relative to the
118 deterministic win–stay/lose–switch (WSLS) expectation (Supplementary Fig. 1b), showed that
119 rats did not simply react to the immediately preceding outcome. Consistent with this history-
120 dependent strategy, logistic regression on recent trial history showed that choice and choice–
121 outcome interaction terms were strongest at short lags and decayed across the preceding
122 trials (Fig. 1e).

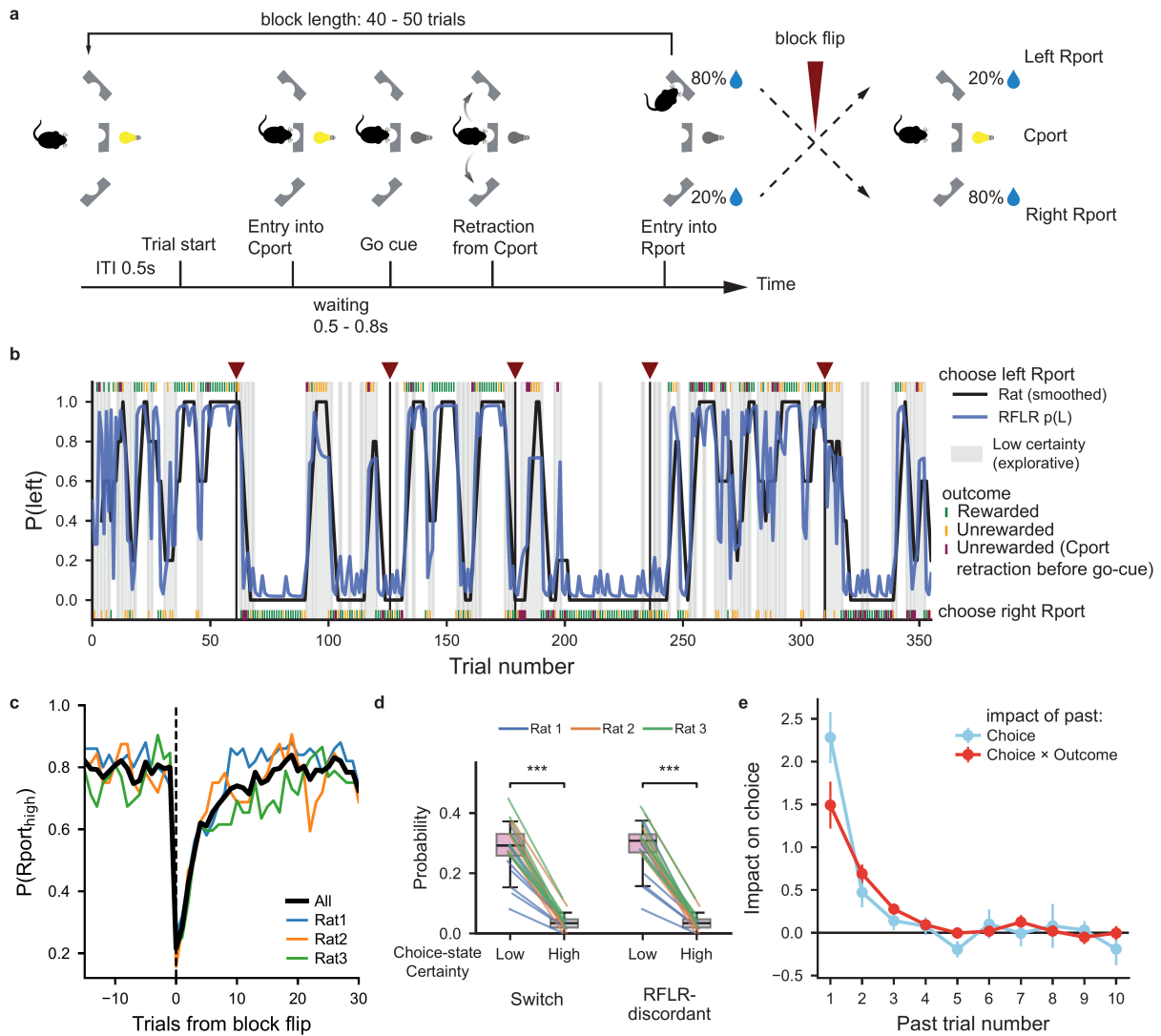


Figure 1 | Rats adapt to unsignaled reversals and increase exploration of alternative choices when model-derived choice-state certainty is low. **a**, Two-armed bandit reversal task. Each trial began with central port (Cport) entry and proceeded to a left or right reward port (Rport). Reward probabilities were asymmetric (80% versus 20%) within blocks of 40-50 trials and reversed without explicit cues. **b**, Representative session. Black trace, smoothed left-choice probability (five-trial window); blue trace, recursively formulated logistic regression (RFLR)-predicted probability of choosing left. Ticks denote trial outcomes: rewarded (green), unrewarded (orange) and unrewarded premature Cport-retraction trials (magenta); triangles mark reversals. Gray shading marks low-certainty trials, defined as the bottom within-session tertile of RFLR choice-state certainty, $|P(\text{left}) - 0.5|$. **c**, Higher-reward-port choice probability aligned to reversals. Colored lines show the mean for each individual rat across blocks, pooled from all recording sessions; black line shows the population mean. Choices shifted toward the newly advantageous port over approximately 10–15 trials. **d**, Session-level behavioral measures in low- and high-certainty states. Left, probability of switching from the immediately preceding port choice. Right, probability of RFLR-discordant choices, defined as choices opposite to the side favored by the RFLR model ($P(\text{left}) > 0.5$ favors left; $P(\text{left}) < 0.5$ favors right). Low and high states were the bottom and top within-session tertiles of choice-state certainty, respectively. Boxes show session distributions; lines connect paired low/high values for individual sessions colored by rat. Significance was assessed using linear mixed-effects models with rat identity as a random effect ($***P < 0.001$; $n = 22$ sessions, 3 rats). **e**, Logistic regression coefficients quantifying the influence of choices and choice–outcome interactions over the preceding ten trials. Both effects were strongest at short lags and decayed across trials. Points and error bars show mean $\pm$ s.e.m. across sessions.

We next used the RFLR model to define model-derived choice-state certainty, $C_t^{\text{state}} = 2|P_t(L) - 0.5|$, where $P_t(L)$ is the model-estimated probability of choosing the left port before the choice on trial t . This measure indexes the polarization of the inferred action policy: values near 0 indicate weak model preference between the two actions, whereas values near 1 indicate strong model preference for one action. It does not, by itself, define whether the animal ultimately chose the model-favored action. We therefore treated model-derived certainty as a state variable and separately quantified whether behavior was switch-like or model-concordant. Within each session, low- and high-certainty states were defined as the bottom and top tertiles of C_t , respectively. Low-certainty states showed both higher switching from the previous port choice and higher RFLR-discordant choice probability than high-certainty states (Fig. 1d). Thus, low model-derived choice-state certainty was associated with exploration-like behavior, whereas high-certainty states were dominated by stable, model-concordant choices.

Additional held-out model comparisons supported using RFLR-derived certainty as a behavioral state variable. RFLR and Q-learning models predicted held-out choices better than simple WSLS-like and bias-only models, and produced highly concordant trial-by-trial certainty estimates (Supplementary Fig. 1a, c). Low-certainty trials were also enriched for Q-value non-greedy choices (Supplementary Fig. 1d). We therefore use low- and high-certainty states throughout the neural analyses to refer to model-derived choice states, not subjective confidence: low certainty denotes weakly polarized action preference with increased switching and model-discordant choices, whereas high certainty denotes strongly polarized, exploitation-like choice states.

Low-certainty states contain weaker yet rapidly alternating mPFC-vIOFC choice representations

We used ultraflexible electrode arrays to record single-unit activity simultaneously from medial prefrontal cortex (mPFC) and ventrolateral orbitofrontal cortex (vIOFC) while rats performed the reversal bandit task (Fig. 2a). This approach provided comparable single-unit yields across repeated sessions in each animal, allowing population-level effects to be tested across 22 sessions from three rats over approximately two weeks. Across sessions, we recorded 67.2 ± 5.1 mPFC neurons and 66.7 ± 3.5 vIOFC neurons per session (mean $\pm$ SEM), with 6–8 sessions per animal.

Single-neuron activity showed state-dependent selectivity during the task. During the pre-action decision window of the Cport-hold period, neurons selective for upcoming choice were more common in high- than low-certainty states in both regions (mPFC: $46.7 \pm 1.9\%$ versus $25.0 \pm 2.5\%$; vIOFC: $32.5 \pm 2.3\%$ versus $17.6 \pm 2.0\%$; mean $\pm$ SEM across sessions; Fig. 2b). Choice selectivity was also more prevalent in mPFC than vIOFC. By contrast, after unrewarded outcomes, similar fractions of neurons distinguished trials followed by a switch from those followed by a stay in the two regions (mPFC: $11.3 \pm 1.3\%$; vIOFC: $10.0 \pm 1.5\%$; Fig. 2b). Thus, single-neuron selectivity reflected both regional specialization and behavioral state, with the clearest certainty dependence appearing in pre-action upcoming-choice coding.

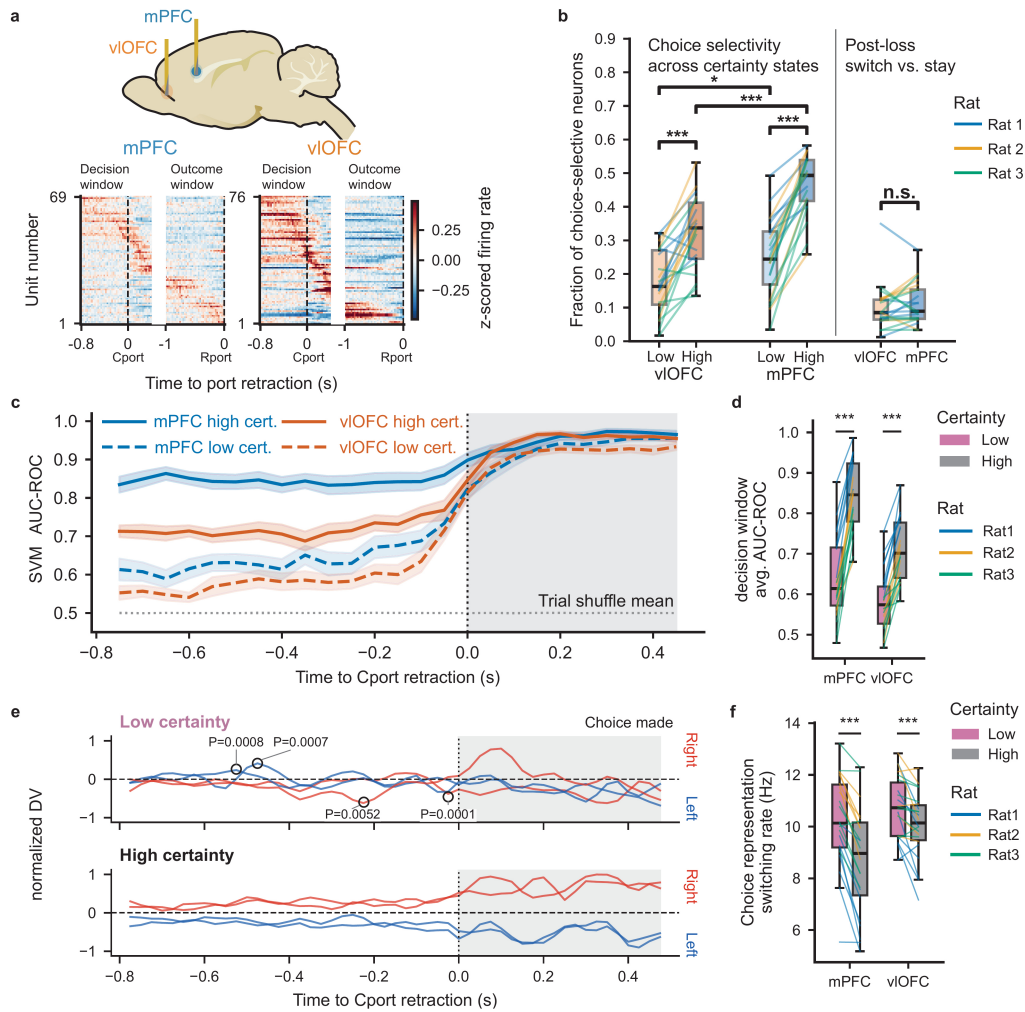


Figure 2 | Neural choice representations stabilize during high-certainty exploitation but rapidly alternate between competing choice signals during low-certainty exploration. **a**, Simultaneous recordings from medial prefrontal cortex (mPFC; prelimbic/infralimbic cortex) and ventrolateral orbitofrontal cortex (vIOFC). Top, recording locations. Bottom, trial-averaged z-scored firing rates from one session; neurons are sorted by peak response time. Dashed lines mark central-port (Cport) and reward-port (Rport) retraction. **b**, Fraction of neurons showing selective firing patterns for upcoming choice during the pre-action decision window (-0.6 to -0.1 s before retraction). High-certainty trials recruited more choice-selective neurons than low-certainty trials in both regions, and higher abundance in mPFC than vIOFC generally across certainty states. Fraction of neurons showing post-loss switch-versus-stay selectivity did not differ between regions (linear mixed-effects models; not significant, $P > 0.05$; * $P < 0.05$; *** $P < 0.001$). Lines connect sessions and are colored by rat. **c**, Time-resolved linear support vector machine decoding of upcoming choice during Cport hold, quantified by area under the receiver operating characteristic curve (AUC-ROC). High- and low-certainty trials are the top and bottom within-session tertiles of recursively formulated logistic regression (RFLR) choice-state certainty, $2|P(\text{left}) - 0.5|$. Shading, standard error of the mean (s.e.m.); gray dotted line, trial-shuffle mean. **d**, Mean decision-window decoding performance, which was higher during high-certainty states in both regions (linear mixed-effects models; *** $P < 0.001$). **e**, Single-trial normalized decision-variable (DV) traces from one vIOFC session; two trials per eventual choice and certainty condition. Positive and negative DV indicate right- and left-choice evidence. Low-certainty examples were selected for the smallest P values for excursions toward the unchosen side; high-certainty examples were selected to minimize excursions. Display smoothing: Gaussian (SD, 20 ms). Circles mark unsmoothed values; labels report two-sided P values from 9,999 choice-label permutations with maximum-statistic correction across all session trials and pre-action bins. **f**, Session-level rates, in hertz (Hz), of switching between choice signals during the pre-action window (-0.8 to 0 s). Switching was faster during low-certainty exploratory states in both regions (linear mixed-effects models; *** $P < 0.001$). Boxes show session distributions; lines connect sessions and are colored by rat.

212 We next asked whether these state-dependent single-neuron effects were reflected at the
213 population level. A linear support vector machine decoded upcoming choice from activity
214 during the central-port hold period in both regions (Fig. 2c). Decoding performance during the
215 pre-action decision window was higher in high- than low-certainty states (-0.6 to -0.1 s before
216 central-port retraction; mPFC AUC–ROC: 0.841 ± 0.019 versus 0.637 ± 0.022 ; vIOFC: 0.712
217 ± 0.018 versus 0.581 ± 0.016 ; mean $\pm$ SEM across sessions; linear mixed-effects models with
218 rat identity as a random effect, both $P < 10^{-20}$; Fig. 2d). Thus, model-derived choice-state
219 certainty was associated with the strength and readability of cortical population choice
220 representations.

221 Lower decoding performance in low-certainty states could reflect temporally variable choice
222 evidence within individual trials. We therefore tracked a time-resolved posterior decision
223 variable (DV) from held-out population activity. Positive and negative DV values indicated
224 momentary evidence for right and left choices, respectively; values near zero indicated weak
225 separation between decoded choices. To illustrate these dynamics, we selected four low-
226 certainty and four high-certainty trials from the same vIOFC session, balancing eventual left
227 and right choices (Fig. 2e). Low-certainty examples were selected for excursions toward the
228 unchosen side with the smallest within-session adjusted P values. Each contained a marked
229 excursion exceeding the session-wide permutation threshold ($P = 0.0001$ – 0.0052). Three
230 remained significant after additional correction across the 22 sessions examined (Methods).
231 High-certainty examples were selected to minimize excursions: three had none beyond the
232 defined DV threshold, and one had a single bin exceeding it. Smoothing was applied only for
233 display; selection and statistical assessment used unsmoothed DV traces.

234 Across sessions, pre-action DV sign-switching rates were higher in low- than high-certainty
235 states in both regions (Fig. 2f). This comparison used session-level summaries of the certainty
236 groups, independently of which example trials were displayed. Additional control analyses
237 supported this interpretation (Supplementary Fig. 2a-d). The low-certainty increase in
238 switching persisted across DV threshold choices and after matching high- and low-certainty
239 trials for mean DV amplitude, arguing against a simple weak-signal explanation. A choice-
240 label shuffle showed that observed low-certainty switching exceeded the shuffled-label null in
241 mPFC, whereas this comparison was weaker in vIOFC. To test whether the low-high contrast
242 itself depended on the observed certainty assignment, we kept decoded DV metrics fixed and
243 permuted low/high certainty labels across trials. The observed low-high switching difference
244 exceeded this certainty-label null in both regions (500 shuffles; linear mixed-effects models,
245 mPFC: $\Delta = 0.692$ Hz, $P = 1.67 \times 10^{-21}$; vIOFC: $\Delta = 0.388$ Hz, $P = 1.96 \times 10^{-6}$).

246 Together, these analyses show that model-derived choice-state certainty was reflected in
247 cortical population dynamics during the Cport-hold period. High-certainty states were
248 associated with stronger and more persistent upcoming-choice representations, whereas low-
249 certainty states retained choice information in a less stable, rapidly alternating form,
250 particularly in mPFC. This population-level state dependence motivated the next analysis, in
251 which we asked whether coordinated mPFC–vIOFC ensembles organize competing pre-
252 action choice representations.

Low-certainty states simultaneously recruit competing choice-selective inter-areal ensembles, whereas high-certainty states stabilize pre-action choice representations

Neurons in mPFC and vIOFC formed structured inter-areal ensembles during the central-port hold period before movement initiation. Because population decoding showed that both regions encoded upcoming choice during this period, we next asked whether these signals were organized by coordinated activity among specific groups of neurons. We identified ensembles with PCA–ICA applied to population activity in the pre-action decision window, defined as -0.8 to 0 s before Cport retraction.^{25,26} Ensemble activation was quantified as a weighted co-activation signal derived from each ensemble's PCA–ICA loading vector, such that higher activation reflects coordinated activity among neurons contributing strongly to that ensemble rather than the firing rate of a single neuron alone.

Choice-selective inter-areal ensembles were identified without assuming a fixed latency of choice coding. We screened 250-ms windows across the Cport-hold period using a maximum-likelihood choice-label shuffle, which selected ensembles with significant time-flexible choice selectivity across animals (Supplementary Fig. 3a). These ensembles typically contained neurons from both regions, but their high-weight members were biased toward mPFC (Fig. 3a). The weight-defined membership was also functionally meaningful: high-weight members showed higher normalized firing activity and stronger ensemble-favoured choice modulation than nonmembers across session-level averages (Supplementary Fig. 3b,c). In an example ensemble, activation increased before the ensemble-favoured upcoming choice and was strongest during high-certainty trials (Fig. 3b). Across choice-selective, certainty-modulated ensembles, activity was stronger for the ensemble-favoured choice during high- than low-certainty states (Fig. 3c; linear mixed-effects model, $P < 0.001$).

We next examined whether low-certainty states were marked only by weaker ensemble activation, or instead by a different organization of competing choice representations. In single trials, low-certainty states showed transient recruitment of ensembles preferring opposite future choices within the same Cport-hold period, whereas high-certainty trials showed more stable dominance of ensembles aligned with the subsequent behavioral choice (Fig. 3d). To quantify this population-state structure, we measured two complementary features. First, we measured alternative-choice ensemble activity, defined as activation of ensembles preferring the action not chosen on that trial. Second, we measured final-state choice congruence, defined as whether the ensemble preference dominating immediately before Cport retraction matched the animal's subsequent port choice. Alternative-choice ensemble activity was stronger in low- than high-certainty trials, whereas final-state choice congruence was stronger in high-certainty trials (Fig. 3e; linear mixed-effects models, both $P < 0.01$).

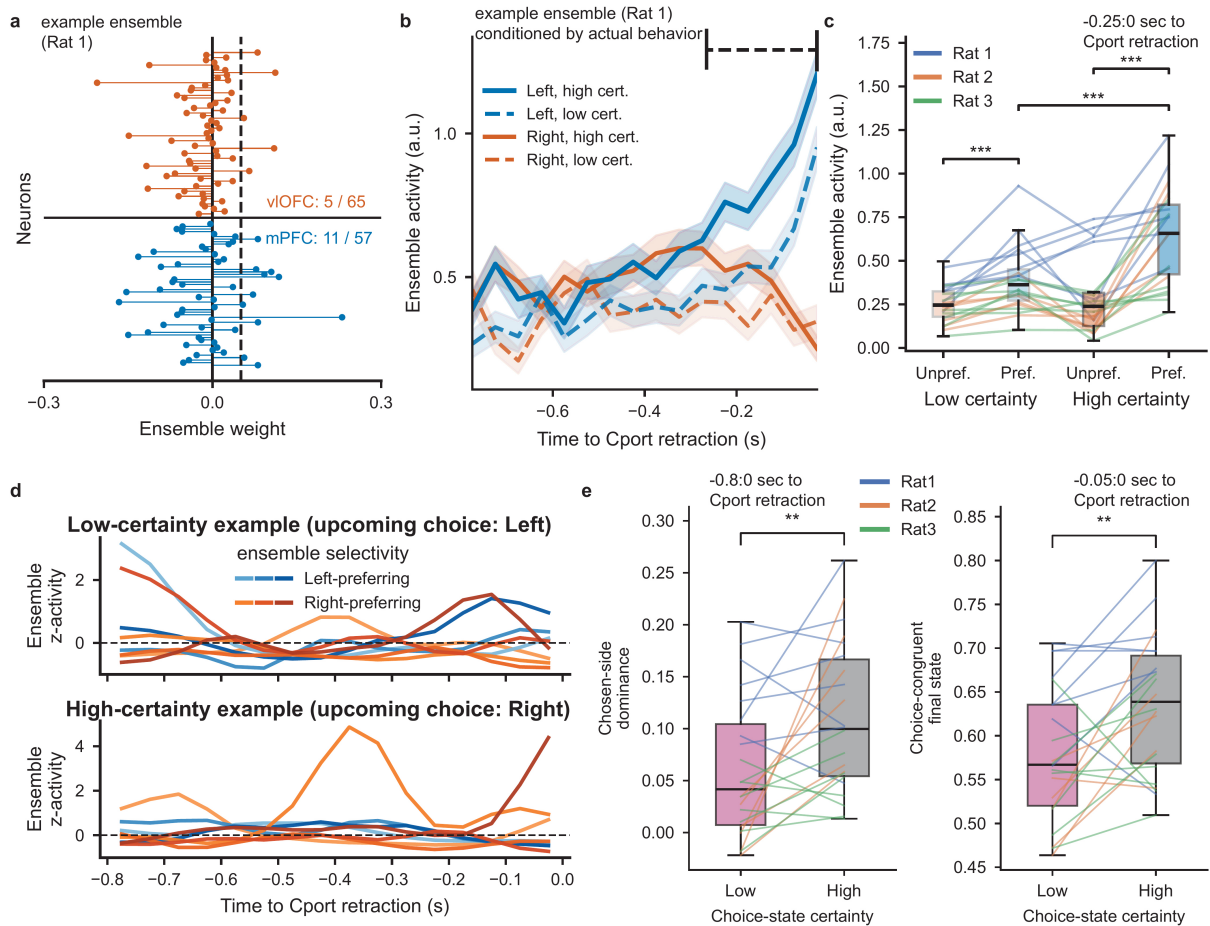


Figure 3 | Low-certainty states recruit competing and alternating choice-selective inter-area ensembles, whereas high-certainty states stabilize pre-action choice representations. Inter-area ensembles spanning medial prefrontal cortex (mPFC) and ventrolateral orbitofrontal cortex (vOFC) were identified during the central-port (Cport) hold period before movement initiation. Low and high certainty were the bottom and top within-session tertiles of recursively formulated logistic regression (RFLR) choice-state certainty. **a**, Weights of mPFC (blue) and vOFC (orange) neurons in an example ensemble. Points represent neurons; the dashed line marks the membership threshold (mean + 1 standard deviation of ensemble weights). **b**, Trial-averaged activity, in arbitrary units (a.u.), of the example ensemble aligned to Cport retraction and grouped by upcoming choice and certainty. Blue and orange denote upcoming left and right choices; dashed lines denote low-certainty trials. Activity was strongest for the ensemble-favored choice under high certainty. Shading indicates s.e.m. across trials. **c**, Activity of choice-selective, certainty-modulated ensembles during opposite-choice and ensemble-favored-choice trials under low and high certainty. Activity was higher for ensemble-favored than opposite choices under low certainty and highest for ensemble-favored choices under high certainty ($***P < 0.001$). **d**, Example single-trial ensemble activity. Blue and orange traces depict activity of left- and right-favored ensembles during example high- and low-certainty trials. Low-certainty trials showed alternating recruitment of ensembles representing alternative choices, whereas high-certainty trials showed stable dominance of ensembles aligned with the eventual choice. **e**, Session-level ensemble dynamics. Left, chosen-side ensemble dominance during the pre-action decision window (-0.8 to 0 s before Cport retraction). Right, choice-congruent activation of the final dominant ensemble state (-0.05 to 0 s before Cport retraction). Both measures were higher under high certainty ($**P < 0.01$). In **c** and **e**, boxes show session distributions and colored lines connect sessions by rat. Significance was assessed using linear mixed-effects models with certainty as a fixed effect and rat identity as a random intercept.

Together, these results indicate that low-certainty states were not simply associated with weaker choice-selective ensemble activity. Instead, low-certainty Cport-hold periods recruited competing ensembles representing alternative future actions, whereas high-certainty periods stabilized ensemble activity aligned with the eventual choice. Thus, model-derived choice-state certainty organized both the composition and temporal stability of inter-areal pre-action choice representations.

vIOFC-preceding mPFC outcome signals accompany switching upon unrewarded choices after high-confidence choices (i.e. surprises)

Negative outcomes prompt animals to re-evaluate whether to continue exploiting the current choice policy or shift toward an alternative. We therefore examined how post-outcome activity evolved across mPFC and vIOFC during unrewarded trials. During reward-port holding, a subset of inter-areal ensembles was preferentially recruited after reward omission and activated at staggered times across the post-outcome window (Fig. 4a). These loss-responsive ensembles were dominated by high-weight vIOFC members, while still retaining inter-areal membership (Fig. 4b). The same pattern was observed in additional animals, where Rat 2 and Rat 3 both showed loss-responsive ensemble traces with vIOFC-dominant membership structure (Supplementary Fig. 4). Thus, reward omission recruited temporally organized vIOFC-weighted inter-areal ensemble cascades rather than a single uniform post-outcome response.

We next asked whether outcome representations in mPFC and vIOFC showed a consistent temporal relationship. We decoded current-trial outcome from population activity in each region and tracked the distance to the decoder decision boundary over time, aligned to reward-port retraction. Negative values corresponded to activity patterns associated with reward omission. In an example session, both regions showed clear outcome-decoding signals, with rewarded and unrewarded trials separating in the post-outcome period (Fig. 4c). For each unrewarded trial, we estimated the temporal offset between regional outcome signals from the center of mass of the mPFC–vIOFC cross-correlation profile (Fig. 4d). Negative lag values indicate vIOFC-preceding dynamics, whereas positive values indicate mPFC-preceding dynamics. This analysis quantifies temporal precedence between regional population signals, not causal directionality.

We then focused on unrewarded trials after model-concordant choices, where the animal had chosen the action favored by the behavioral model. In high-certainty trials followed by a port switch, outcome signals showed stronger vIOFC-preceding dynamics than in high-certainty trials followed by a stay (Fig. 4e). High-certainty switch trials also showed stronger vIOFC-preceding dynamics than low-certainty switch trials. By contrast, low-certainty stay and switch trials showed similar lag structure. Thus, vIOFC-preceding outcome dynamics were most evident when an unrewarded outcome followed exploitation of a strongly favored policy and preceded a subsequent behavioral switch.

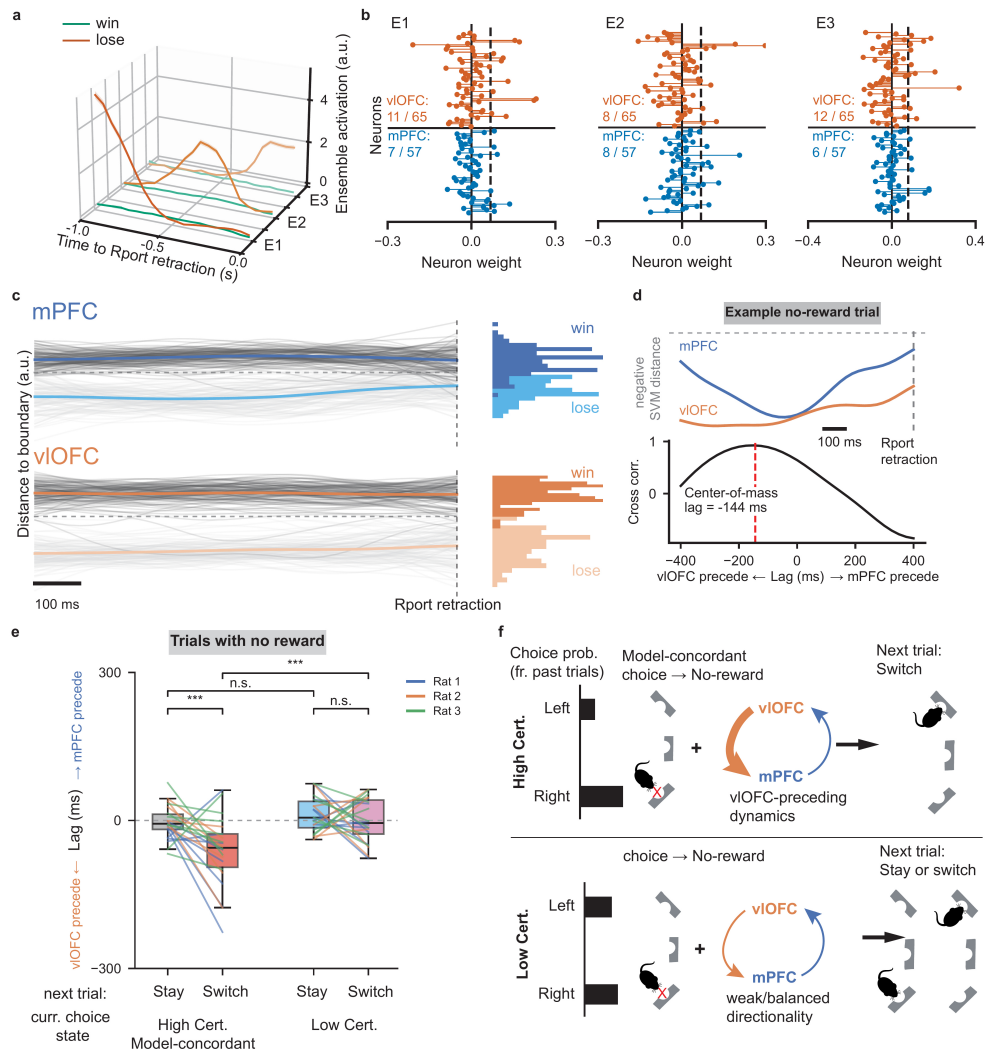


Figure 4 | During losses purely loss-responsive vIOFC-dominated ensembles emerge, and choice-switching after unrewarded outcomes upon high-confidence choices arises following vIOFC → mPFC dominated directional inter-areal dynamics.

a, Loss-responsive inter-areal ensemble dynamics during reward-port (Rport) holding. Example ensemble activations (E1–E3), aligned to Rport retraction (time 0), show preferential recruitment after unrewarded trials and sequential activation across the post-outcome period. **b**, Neuron weights contributing to the outcome-related ensemble axis in each epoch. Orange and blue denote vIOFC and mPFC neurons, respectively; numbers indicate neurons with positive weights out of recorded neurons in each region. **c**, Outcome-decoding signals in mPFC and vIOFC from an example session. Each trace shows the distance to a linear support vector machine decision boundary for current-trial outcome, aligned to Rport retraction. Grey traces show individual trials. Thick coloured traces show condition means for rewarded and unrewarded trials within each region; darker colours denote rewarded trials and lighter colours denote unrewarded trials. Right histograms summarize end-of-trial decoding values. **d**, Example unrewarded trial showing the temporal relationship between mPFC and vIOFC outcome-decoding signals. The red dashed line marks the center-of-mass lag of the cross-correlation profile; negative values indicate vIOFC-preceding dynamics and positive values indicate mPFC-preceding dynamics. **e**, Lag estimates for unrewarded trials after model-concordant choices, grouped by current choice-state certainty and next-trial action. Model-concordant choices are choices matching the action assigned higher probability by the behavioral model. Boxes show session distributions; paired lines connect session means and are coloured by rat. Negative values indicate vIOFC-preceding dynamics, and positive values indicate mPFC-preceding dynamics. Comparisons were tested with linear mixed-effects models with rat identity as a random effect; n.s., not significant; *** $P < 0.001$. **f**, Schematic summary: after high-certainty model-concordant unrewarded choices, vIOFC-preceding dynamics are stronger before subsequent switching, whereas low-certainty unrewarded choices showed weaker or more balanced temporal offsets between regional outcome signals.

378 Together, these results show that reward omission recruits vOFC-dominated ensemble
379 cascades and temporally asymmetric mPFC–vOFC outcome signals. In high-certainty model-
380 concordant trials, vOFC-preceding dynamics were associated with subsequent switching,
381 linking post-outcome inter-areal population dynamics to the updating of behavioral policy after
382 failed exploitation (Fig. 4f).

383 Discussion

384 Adaptive decisions require neural circuits to preserve useful choice policies despite noisy
385 feedback while remaining able to revise those policies when environmental contingencies
386 change. We identified certainty-dependent population dynamics that may help balance choice
387 persistence with behavioral flexibility. During the Cport-hold period before action execution,
388 high-certainty choices were represented strongly and stably in both mPFC and vOFC,
389 whereas low-certainty choices were preceded by rapid alternations between representations
390 of the available actions. Certainty-dependent choice encoding was also evident at the
391 ensemble level: choice-selective, certainty-modulated inter-areal ensembles were dominated
392 by mPFC neurons and represented the eventual choice most strongly under high certainty.
393 Following reward omission, sequential loss-selective ensembles dominated by vOFC neurons
394 emerged throughout the post-outcome period. After losses following high-confidence choices
395 (i.e. surprises), vOFC outcome representations preceded those in mPFC on trials followed by
396 a choice switch. A technical feature of this study was the ability of ultra-flexible tentacle
397 electrodes to maintain comparable single-unit yields in both regions across 6–8 recording
398 sessions spanning approximately two weeks, allowing these population signatures to be
399 evaluated repeatedly across days.²³ For ensemble detection, cross-session unit matching
400 additionally enabled pooling activity from the same neuronal population across sessions within
401 each animal. This provided sufficient observations to estimate coactivation structure spanning
402 mPFC and vOFC.

403 Flexible decision-making requires a favored action to remain sufficiently stable to guide
404 behavior while alternative actions remain available when support for that choice weakens. Our
405 results suggest that model-derived choice-state certainty was associated with the temporal
406 organization, as well as the strength of choice representations. Under high certainty, mPFC
407 and vOFC populations maintained a stable representation of the eventual choice, which may
408 protect a well-supported preference from stochastic negative feedback. Under low certainty,
409 decoded activity repeatedly alternated between the subsequently chosen and alternative
410 actions. Such alternation may keep competing actions available for comparison and allow new
411 evidence to redirect behavior. Previous OFC recordings similarly revealed alternation between
412 population states representing the values of the available options during subjective
413 decisions.²⁷ Our findings relate the stability of alternative-action representations to model-
414 derived choice-state certainty and extend this phenomenon to simultaneously recorded mPFC
415 –vOFC populations and interareal ensembles. This interpretation is compatible with active-
416 inference models linking policy selection to precision and epistemic value,²⁸ and with
417 information-seeking accounts of exploration.²⁹ This comparison is conceptual: low choice-state
418 certainty does not establish low policy precision, and increased switching does not
419 demonstrate directed information seeking. The results are also compatible with affordance
420 competition as a framework in which representations of potential actions are specified
421 concurrently and compete for selection.³⁰ In our data, model-derived certainty distinguished
422 conditions under which one action representation stabilized relative to its alternative. Evidence

423 that beta synchrony organizes rule-specific prefrontal ensembles and that beta-band signals
424 distinguish categorical decisions motivates the hypothesis that beta activity could contribute
425 to these representational transitions.^{31,32} However, our measure counted sign transitions in
426 decoded decision-variable traces rather than oscillatory activity. Moreover, because a
427 complete alternation cycle contains two opposite-direction transitions, the sign-transition rate
428 should not be interpreted as an equivalent oscillation frequency. Determining whether beta
429 phase or spike-field coupling organizes these transitions will require direct
430 electrophysiological analyses and causal tests.

431 A flexible decision system must integrate specialized regional computations while forming a
432 coherent action. We found that choice-selective, certainty-modulated ensembles generally
433 included neurons from both mPFC and vOFC, but their strongest contributors were biased
434 toward mPFC and represented the eventual action most strongly under high certainty. This
435 organization should be interpreted against evidence for regional differentiation. Lesion studies
436 in marmosets dissociated lateral prefrontal contributions to attentional shifting from OFC
437 contributions to affective reversal,³³ while comparative anatomy identified orbital and medial
438 prefrontal cortex as distinct but interconnected networks.³⁴ Rodent mPFC contributes to action
439 selection, memory-guided decisions and stable representations of decision variables,^{22,35,36}
440 whereas OFC has been implicated in expected-outcome and latent task-state
441 representations.^{6,8} More specifically, medial OFC supports outcome retrieval when task states
442 are partially observable.¹⁹ Reward-schedule uncertainty can also modulate OFC choice and
443 outcome coding,³⁷ and OFC confidence signals can generalize across behaviors and sensory
444 modalities.³⁸ These constructs and OFC subregions differ from the trial-wise polarization of
445 action preference and the vOFC recordings examined here. At the population level, our
446 findings are compatible with frameworks treating neuronal ensembles as functional units of
447 computation^{39–41} and proposing that interareal interactions involve selected population activity
448 patterns.^{11,12} Local mPFC assemblies show behavior-dependent temporal organization,⁴² and
449 learning can reorganize theta coordination and spike timing across hippocampal–prefrontal
450 networks.⁴³ Other hippocampal–frontal studies show that awake hippocampal ripple events
451 selectively coordinate prefrontal ensembles,⁴⁴ that hippocampal input causally supports
452 prefrontal spatial coding,⁴⁵ and that communication subspaces align with spatial-memory
453 behavior.⁴⁶ Coordinated population activity has also been observed between dorsal CA1 and
454 medial OFC during flexible goal learning.⁴⁷ These studies provide comparative precedents for
455 selective interareal organization; they do not demonstrate a direct functional pathway between
456 the mPFC and vOFC populations recorded here. Our findings extend this framework to two
457 frontal regions by showing that the composition and momentary dominance of their shared
458 ensembles vary with model-derived choice-state certainty. Because regional dominance
459 within an ensemble does not establish causal control, pathway-specific perturbations will be
460 needed to determine how mPFC, vOFC and possible common inputs generate these
461 dynamics.

462 The behavioral meaning of an unrewarded outcome depends on the decision state in which it
463 occurs. When choice-state certainty is already low, a loss is compatible with an unstable action
464 preference; when certainty is high, the same loss conflicts with a strongly favored action and
465 may provide evidence that the current policy should be reconsidered. In our data, losses
466 recruited sequential, vOFC-dominated ensembles, while vOFC-preceding outcome signals
467 were associated with subsequent switching specifically after high-certainty choices. Several
468 distinct OFC functions provide context for these findings. OFC has been proposed to represent
469 latent task states,⁸ medial OFC supports retrieval of expected outcomes when task states are

470 partially observable,¹⁹ and OFC contributes to updating beliefs used for state inference.⁴⁸
471 Studies of lateral OFC identify additional, but distinct, relationships with adaptive behavior.
472 Reward-schedule uncertainty modulates OFC encoding of choices and outcomes,³⁷ whereas
473 inhibition of lateral OFC eliminates a risky-win bias while sparing spatial win-stay and lose-
474 shift effects.⁴⁹ Lateral OFC is also required for estimating subjective preference during initial,
475 but not established, economic choices,⁵⁰ and can guide value-dependent remapping of
476 sensory cortical representations.⁵¹ These findings concern different forms of uncertainty,
477 preference and behavioral updating; they do not directly explain the spatial post-loss switching
478 or model-derived certainty effects measured here. The interpretation of outcomes by OFC
479 also depends on its inputs: basolateral amygdala projections to OFC causally enable cue-
480 triggered reward expectations.⁵² Pathway-specific chemogenetic inhibition in rats further
481 revealed dissociable contributions of reciprocal medial OFC–prelimbic projections to risky
482 choice and feedback sensitivity.⁵³ This provides a causal precedent for frontal interactions
483 during probabilistic discounting, while differing from the vOFC recording site and reversal-
484 learning task examined here. More direct comparative precedents for frontal coordination
485 show that decision-making dynamically shapes communication within ventral frontal cortex,¹⁵
486 that reward-related choices affect the relative timing and flow of information across lateral
487 prefrontal areas,²⁰ and that flexible sensorimotor decisions involve dynamically organized
488 cortical information flow.⁵⁴ Against this background, our findings identify a context-dependent
489 temporal association between mPFC and vOFC outcome representations. However, cross-
490 correlation lag establishes temporal precedence rather than direct information transfer or
491 causality. The vOFC-preceding outcome signals could reflect direct vOFC influence on
492 mPFC, a polysynaptic pathway or common input expressed at different latencies.

493 Our conclusions are bounded by two principal limitations. First, certainty was inferred from
494 choice history as the distance from model-predicted indifference. This measure captures the
495 strength of the animal's current choice preference but does not distinguish subjective
496 confidence, expected uncertainty, environmental volatility, or surprise.^{2,55–57} Second, the
497 ensemble and temporal-offset analyses are observational. Although cross-correlation lag
498 establishes temporal precedence, it cannot determine whether vOFC directly influences
499 mPFC or whether the observed relationship arises through polysynaptic pathways or common
500 input. Behavioral manipulations that dissociate these latent variables, together with closed-
501 loop ensemble perturbations and pathway-specific circuit manipulations, will be necessary to
502 establish the causal mechanisms underlying the observed dynamics. Within these constraints,
503 our findings identify complementary population processes through which distributed prefrontal
504 circuits may balance decision stability and flexibility. High-certainty choices were supported
505 by stable, mPFC-dominated inter-areal ensembles, whereas low-certainty states permitted
506 rapid alternation between competing action representations. When an established high-
507 certainty policy failed, sequential vOFC-dominated loss ensembles and stronger vOFC-
508 preceding mPFC temporal offsets accompanied subsequent behavioral switching. Flexible
509 decision-making may therefore depend not only on which variables mPFC and OFC represent,
510 but also on how certainty reorganizes ensemble recruitment, temporal stability and inter-areal
511 timing of their population activity.

512 Methods

513 Animals

514 Female Long Evans rats were used in this study. All rats were purchased from Charles River
515 Laboratories (France). Rats were group-housed on a reversed 12 h light/dark cycle. Training
516 and recording were performed during the dark phase. Food was available *ad libitum*. At the
517 start of behavioral training, rats weighed approximately 200 g. Rats had scheduled access to
518 water during behavioral training and recording days. Rats obtained water during task
519 performance and received supplementary water after training. Water intake and body weight
520 were monitored daily to ensure that body weight remained above 90% of baseline during
521 behavioral training. All experimental and surgical procedures involving animals were approved
522 by the local veterinary authorities of Canton Zurich, Switzerland, and were carried out in
523 accordance with the guidelines published in the European Communities Council Directives
524 2010/63/EU.

525 Behavioral apparatus

526 Behavioral training and recordings took place in a custom-built three-port operant training
527 chamber. Each nose port contained an infrared beam detector to detect nose pokes and a
528 yellow LED. Both the left and right reward ports were equipped with a licking spout for reward
529 delivery. The entire behavioral setup was enclosed in a sound-attenuating box (Coulbourn
530 Instruments, PA, USA). Inside the box, an infrared lightboard was used to provide illumination
531 for video recording. A video camera mounted on the ceiling of the box recorded behavior. The
532 training pipeline was implemented using a custom Python script, which controlled the
533 experimental parameters and managed the video recording.

534 Two-armed bandit reversal task

535 Before the two-armed bandit reversal task, rats underwent pre-training, during which the rats
536 learned to initiate the task by poking and remaining in the central nose port until the go cue. If
537 the rats retracted from the central port before the go cue, the trial was terminated and the rats
538 would not receive a reward from either reward port. The required waiting duration was
539 gradually increased from 0.05 to 1.5 s across days. The duration was increased when a rat
540 waited for the go cue in more than 90% of trials. During this pre-training phase, the rats were
541 free to collect a reward from either reward port. If a rat displayed a strong bias towards either
542 reward port at the early training stage, forced blocks were implemented to correct the side
543 bias, in which only the non-biased reward port provided a reward. Rats proceeded to electrode
544 implantation after learning to alternate between the two reward ports.

545 After recovering from electrode implantation, rats were connected to the recording system and
546 habituated in the training chamber for two short sessions (20 minutes each). During these
547 sessions, the rats performed the same paradigm as in the pre-training phase. This step helped
548 the rats to habituate to the cables connecting the recording system. Following the habituation,
549 the rats were introduced to the probabilistic two-armed bandit reversal task. The probabilistic
550 two-armed bandit reversal task followed the same trial structure as pre-training, except that
551 reward delivery became probabilistic. In each trial, the rewards were delivered stochastically,
552 with no explicit sensory information to indicate the better choice. In a block of trials containing
553 approximately 40-50 trials, one reward port had an 80% chance to trigger a reward, while the
554 other had a 20% chance. Reward probabilities were constant within a block of trials and
555 switched in the subsequent block of trials without explicit sensory information to signal the

change. Rats could identify the higher-probability port only through trial-by-trial feedback. Each session typically contained 5-8 blocks, with the arrangement of blocks randomly initiated at the beginning of each training session. Electrophysiological data and video of the behavior were acquired simultaneously when rats performed the task.

Surgery

Surgeries were performed under aseptic conditions. A rat was anesthetized using isoflurane (4% for induction and 1.8–2.0% for maintenance). Analgesia (Meloxicam, 2 mg/kg body weight, s.c.) was applied prior to surgery. Scalp hair was removed with a trimmer and depilatory gel. Local analgesia was applied subcutaneously (Lidocaine, max 10 mg/kg body weight and max 2 ml/kg body weight) to the scalp. The exposed skin was cleaned sequentially with saline, betadine and 70% ethanol. The rat was then transferred to a surgical table and head-fixed on a stereotaxic apparatus (KOPF Instruments). Body temperature was maintained using a closed-loop small animal heating pad (World precision instruments) placed under the rat. An incision was made in the skin along the midline of the skull. Connective tissue was carefully removed from the skull. The skull was leveled using the Bregma and Lambda as landmarks. Five stainless-steel screws were used to secure the head cap. A head-screw was secured above the cerebellum and used as the ground for the electrophysiological recording system. Two small boreholes with a diameter of 1 millimeter were created above mPFC and IOFC. The two boreholes were temporarily sealed using kwik-cast silicone sealant (World precision instrument) to reduce potential contamination of the brain tissue from the following procedures. Then a layer of dental cement (Tetric EvoFlow, Ivoclar Vivadent) was applied to cover the exposed skull except the area around the two boreholes. A 3D printed head-cap was secured on top of the skull using dental cement. Kwik-Cast was removed from the boreholes, and the dura was removed to expose the brain. Electrode arrays were then inserted into the targeted brain areas using a micromanipulator with a speed of 20 $\mu\text{m/s}$, targeting the following coordinates: mPFC, AP +2.0 mm, ML 0.6 mm, DV 3.8 mm below the brain surface; vIOFC, AP +3.8 mm, ML 2.5 mm, DV 4.1 mm below the brain surface. The distance between two electrode shanks in the same brain area was around 300 μm . After all the electrode arrays were inserted, the boreholes were filled with Kwik-cast silicone sealant. The excess ribbon cable and PCB connector were embedded and secured in the 3D-printed head cap with dental cement. After surgery, the rats were single-housed in a clean cage for at least 3 days. During this period, the rats were carefully monitored and analgesic was administered daily (2 mg/kg body weight, in oral suspension). Prior to recording, the rats were observed to ensure that they had recovered fully and behaved normally.

Histology

To perform electrical lesioning, rats were deeply anesthetized using isoflurane. Electrical current pulse ($\pm 5 \mu\text{A}$, 2 seconds) was applied to the selected electrode pads. Following lesioning, rats were deeply anesthetized using a mixture of Ketamine (200 mg/kg) and Xylazine (20 mg/kg) and transcardially perfused with 0.1 M phosphate buffered saline (PBS) followed by 4% paraformaldehyde (PFA) solution in PBS. The brains were then extracted and post-fixed in 4% PFA. Brain tissue was sliced into 100- μm thick sections. The slices were stained using NeuN (Invitrogen Antibodies) to visualize the cell layer structures. Electrode locations were estimated from lesion sites and probe tracks.

599 **Recording system**

600 *In vivo* electrophysiological recordings were performed using ZIF-CLIP® (Tucker Davis
601 Technologies) digital headstages and Intan 512 channel RHD Recording System (Intan
602 Technologies). Extracellular signal was digitized using two 64-channel ZIF-CLIP® digital
603 headstages. Raw neural signals were acquired at 20 kHz using the Intan 512 channel RHD
604 Recording System. To timestamp behavioral events, a USB data acquisition card (DAQ
605 USB8451, National Instruments) sent TTL pulses to the digital input channels of the Intan
606 recording controller, synchronizing behavioral events with electrophysiological recordings. A
607 custom-built commutator was used to connect the headstages to the recording system,
608 allowing the animals to freely move while keeping the cables untangled during prolonged
609 recording sessions.

610 **Data preprocessing and spike sorting**

611 To track single units across recording sessions, we used an approach described by
612 Schoonover *et al.*⁵⁸ The method involved concatenating all recording sessions of interest into
613 a single file for spike detection and spike sorting. To reduce file size, only task-related data
614 were extracted for preprocessing. Specifically, for each trial, raw data were epoched from 2
615 seconds before trial initiation and ending at 4 seconds after outcome was revealed. Then the
616 epoched data were concatenated into a single large file for further processing.

617 The concatenated data were processed for spike detection and sorting using an open-source
618 software package JRCLUST⁵⁹. Raw data were band-pass filtered from 300 to 5,000 Hz. Spike
619 detection was performed using a threshold computed based on the noise level of individual
620 channels. The detection threshold was defined as:

621 $\text{Threshold} = \text{qqFactor} \times \frac{\text{med}(\text{abs}(x - \text{med}(x)))}{0.6745}$, which was calculated as the qqFactor

622 multiplied by the median of the absolute deviation from the median of the raw data of the
623 channel (x), divided by 0.6745. We used the qqFactor = 6.5 in this study. This threshold was
624 higher than conventional values (typically 4–5) because we prioritized stable, high-amplitude
625 single units, at the cost of excluding lower-amplitude units. The resulting waveforms were
626 aligned to their peak. Waveforms were then clustered using principal component analysis and
627 a density-peaks search algorithm.

628 The clustering results were inspected and manually curated using the graphical user interface
629 (GUI) of JRCLUST. Specifically, we discarded noise clusters and merged or split remaining
630 clusters based on waveforms, amplitudes, principal component features, inter-spike-interval
631 (ISI) features, autocorrelograms, and cross-correlograms. After curation, the resulting spike
632 data were split into individual recording sessions for further analysis. Only the units that met
633 the following criteria were included: (1) signal-to-noise ratio was greater than 6; (2) mean firing
634 rate across all trials was greater than 1; (3) fewer than 2% of interspike intervals (ISIs) violated
635 the refractory period ($\text{ISI} < 1.5$ ms); and (4) unit waveform similarity was greater than 0.93.

636 Unit waveform similarity was defined from session-wise mean waveforms. First, we computed
637 the mean waveforms of a unit separately for individual sessions. Next, we calculated
638 Pearson's correlation coefficient between the mean waveforms from any two recording
639 sessions for the same unit. The mean of these Pearson correlation coefficients was used as
640 the waveform-similarity measure. Only units with a mean Pearson's coefficient greater than
641 0.93 were considered stable and included for further analysis.

Behavioral modeling

Recursively formulated logistic regression (RFLR). Choice behavior was modeled with a recursively formulated logistic regression model fitted to each animal's behavioral sessions.¹⁶ In the trial tables, choices were encoded as left = 0 and right = 1; for model updating, choices were converted to a signed action variable a_t , with $a_t = +1$ for left choices and $a_t = -1$ for right choices. Outcomes were encoded as $r_t = 1$ for reward and $r_t = 0$ for no reward. The model maintained a signed evidence variable E_t that integrated rewarded choice history with exponential decay. After each trial, evidence was updated as

$$E_{t+1} = \gamma E_t + \beta r_t a_t,$$

where β is the evidence-weighting parameter and $\gamma = \exp(-1/\tau)$ is the decay factor determined by the memory timescale τ . Thus, rewarded choices contributed signed evidence for the chosen side, whereas unrewarded choices did not add new signed evidence but allowed prior evidence to decay. For trials after the first trial, the probability of choosing the left port was

$$P_t(L) = \sigma(E_t + \alpha a_{t-1}),$$

where α captures choice persistence and σ is the logistic function. The first-trial choice probability was initialized at 0.5. Model parameters α , β , and τ were estimated by minimizing the negative log-likelihood of the observed choices. Trial-by-trial RFLR predictions were used to visualize evolving choice preference in representative sessions and to define model-derived choice-state certainty (Fig. 1b).

Model-derived choice-state certainty. The fitted RFLR model provided a trial-by-trial predicted left-choice probability, $P_t(L)$. We defined model-derived choice-state certainty as

$$C_t = 2|P_t(L) - 0.5|,$$

which ranges from 0 to 1. Values near 0 indicate that the model assigned similar probabilities to the two actions, whereas values near 1 indicate a strongly polarized model-derived choice state. Low- and high-certainty trials were defined within each session as the bottom and top tertiles of C_t , respectively. These labels therefore describe the strength of the model-derived choice state, not subjective confidence. Low-certainty periods are illustrated in the representative session in Fig. 1b, and low- versus high-certainty behavioral comparisons are shown in Fig. 1d.

Behavioral characterization by certainty state. To characterize behavior across certainty states, valid trials were mapped back to the full trial table. A switch was defined as a current choice that differed from the immediately preceding choice in the full behavioral sequence; the first trial of a session was excluded from this analysis. We also quantified RFLR-discordant choices, defined as choices opposite to the action favored by the model: left was favored when $P_t(L) > 0.5$, and right was favored when $P_t(L) < 0.5$. Trials with exactly balanced model predictions were excluded from this discordance measure. Switching probability and RFLR-discordant choice probability were summarized for low- and high-certainty trials within each session and compared with linear mixed-effects models using certainty state as a fixed effect and rat identity as a random intercept (Fig. 1d).

Trial-history and win-stay/lose-switch characterization. To visualize recent trial-history effects, we fitted a conventional lagged logistic regression containing choice, outcome, and choice-

by-outcome terms from the preceding ten trials. This analysis was used to summarize the temporal weighting of recent behavioral history (Fig. 1e). Simple win-stay/lose-switch tendencies were quantified from consecutive trial transitions by measuring the probability of staying after rewarded trials, switching after unrewarded trials, and following the win-stay/lose-switch rule overall. These descriptive analyses tested whether behavior could be explained by a simple reaction to the immediately preceding outcome and are summarized in Supplementary Fig. 1b.

Model comparison. To compare predictive models, we fitted seven models to each animal's behavioral sessions: a bias-only model, a bias-plus-perseveration model, two win-stay/lose-switch models, the RFLR model, a Q-learning model, and a hidden-context model. The bias-only model estimated a constant right-choice probability. The bias-plus-perseveration model added a term for the previous action, such that recent right or left choices shifted the current right-choice logit. 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.

The first win-stay/lose-switch model implemented a deterministic transition rule with a lapse parameter. After rewarded trials, the target action was to repeat the previous choice; after unrewarded trials, the target action was to switch. The model assigned probability $1 - \lambda$ to the target action and λ to the opposite action, with λ constrained between 0 and 0.5. The second win-stay/lose-switch model allowed the probability of staying after reward and switching after no reward to vary independently. Both win-stay/lose-switch models included a separate initial right-choice probability for the first trial of each session.

The Q-learning model maintained separate action values for left and right choices, initialized to 0.5 at the start of each session. On each trial, right-choice probability was given by

$$P_t(R) = \sigma[\beta_Q(Q_t(R) - Q_t(L)) + b + \kappa u_{t-1}],$$

where β_Q is the inverse-temperature parameter, b is a side-bias term, κ captures choice perseveration, and u_{t-1} indicates the previous action. After each choice, the chosen action value was updated toward the observed outcome using separate learning rates for rewarded and unrewarded trials. The unchosen value was multiplied by $1 - \phi$, where ϕ is a forgetting parameter.

Models were fitted separately for each animal. Predictive performance was assessed with leave-one-session-out cross-validation: models were fitted on all but one session from an animal and evaluated on the held-out session. Held-out log loss was used as the main predictive-performance metric; Brier score, accuracy, and full-fit AIC/BIC were also computed as secondary summaries. Cross-validated model performance is shown in Supplementary Fig. 1a.

Validation of RFLR-derived certainty. To test whether RFLR-derived certainty reflected a broader model-derived behavioral state rather than an idiosyncrasy of one model, we compared RFLR certainty with Q-learning certainty across sessions. Q-learning certainty was defined analogously as $2|P_t^Q(L) - 0.5|$. Session-level concordance between RFLR and Q-learning certainty estimates was quantified with Spearman correlation (Supplementary Fig. 1c). We also asked whether RFLR low-certainty states were enriched for non-greedy choices under the Q-learning model. A Q-learning non-greedy choice was defined as a choice opposite to the action with the higher Q value on that trial. Non-greedy choice probability was

summarized across within-session tertiles of RFLR certainty and compared with linear mixed-effects models using rat identity as a random effect (Supplementary Fig. 1d).

Single neuron characterization

Trial-averaged peri-event time histogram (PETH). To visualize the firing patterns of individual neurons in an example recording session, trial-averaged peri-event time histograms (PETHs) were generated separately for mPFC and vOFC units. For each valid trial, two task epochs were extracted and concatenated: (1) the Cport epoch, spanning 0.8 s before to 0.4 s after Cport retraction, aligned to Cport retraction; and (2) the reward-port epoch, spanning 1.0 s before to 0 s at reward-port retraction, aligned to reward-port retraction. These two epochs were displayed adjacent to one another with a small visual gap to indicate that they were aligned to distinct behavioral events.

Spike activity was binned using 25 ms bins. For each neuron, binned spike counts were z-scored within session across all valid trial-bin observations from the concatenated Cport and Rport epochs. No condition-based trial selection was applied; all valid trials were included regardless of choice, outcome, or certainty level. The z-scored binned activity was then averaged across trials to obtain one mean activity profile per neuron. For heatmap visualization, neurons were ordered according to the time bin at which their trial-averaged z-scored activity reached its peak across the concatenated Cport and Rport epochs. Thus, neurons with the earliest peak activity appeared at the top of the heatmap.

Single neuron selectivity analysis. Single-unit selectivity was assessed during the pre-action Cport-hold period. For each session, spike activity was aligned to Cport retraction and spikes were counted in a fixed window from 0.6 s to 0.1 s before Cport retraction. Spike counts in this window were used directly for selectivity testing.

Upcoming-choice selectivity was tested separately for low- and high-certainty trials. Current-trial choice labels were extracted from the matched full trial table and normalized to binary left/right labels. Trial certainty was defined as choice probability's absolute distance to 0.5. Within each session, trials were split by certainty tertiles; the bottom tertile was classified as low certainty and the top tertile as high certainty. The middle tertile was not included in the low- versus high-certainty selectivity comparison.

For each neuron, choice selectivity was tested within each certainty group using a two-sided label-permutation test.⁶⁰ The test statistic was the difference in mean spike count between the two choice groups. Choice labels were randomly permuted 5,000 times to generate a null distribution, and the permutation p-value was computed as the proportion of shuffled mean differences whose absolute value was at least as large as the observed absolute mean difference, with a standard +1 correction. Neurons were tested only when both choice groups contained at least eight trials. A neuron was classified as choice-selective if the permutation test yielded $P < 0.05$.

The same pre-action spike-count window was also used to assess lose-switch selectivity. Trials were included when the previous trial outcome was a loss. These trials were labeled according to whether the animal switched choice on the current trial relative to the previous trial or stayed with the same choice. For each neuron, lose-switch selectivity was tested using the same two-sided label-permutation procedure, with a minimum of five trials per group. A neuron was classified as lose-switch selective if $P < 0.05$.

For each session and brain region, the number of tested neurons, number of selective neurons, and fraction of selective neurons were summarized separately for mPFC and vOFC.

Choice-selectivity summaries were generated separately for low- and high-certainty trials, whereas lose-switch selectivity was summarized across all eligible previous-loss trials.

Decoding of choice and outcome signals

We quantified population-level choice and outcome representations with linear decoders applied separately to mPFC and vOFC activity. Analyses were aligned either to central-port retraction for pre-action upcoming-choice decoding or to reward-port retraction for post-outcome decoding.

Time-resolved decoding of upcoming choice. To decode upcoming left versus right choice during the Cport-hold period, spikes were counted in 100-ms windows stepped every 50 ms from -0.8 to 0.5 s relative to Cport retraction. For each session, region, time bin and certainty state, population activity was represented as a trial-by-neuron spike-count matrix. Low- and high-certainty trials were defined from within-session tertiles of the RFLR-derived choice-state certainty measure. To make decoding performance comparable between mPFC and vOFC, the region with more recorded neurons was randomly subsampled to match the region with fewer neurons, and decoding performance was averaged across subsampling repeats.

Linear support vector machine classifiers were trained separately at each time bin to decode upcoming choice. Performance was evaluated with stratified five-fold cross-validation and quantified by AUC-ROC and classification accuracy. Chance-level decoding was estimated by repeating the decoding procedure after shuffling trial labels 1,000 times. Time-resolved AUC-ROC traces and trial-shuffle nulls are shown in Fig. 2c. For session-level summaries, decoding performance was averaged within the pre-action decision window from -0.6 to -0.1 s relative to Cport retraction and compared between low- and high-certainty states with linear mixed-effects models (Fig. 2d).

Posterior decision variable for within-trial choice-representation stability. To examine whether lower decoding performance in low-certainty states reflected loss of choice information or rapid alternation between competing choice representations, we computed a posterior decision variable from single-trial population activity. For this analysis, spikes were counted in 50-ms windows stepped every 25 ms from -0.8 to 0.5 s relative to Cport retraction. Within each cross-validation fold, spike counts were z-scored using the training trials and applied to held-out trials. For each time bin, we computed choice-specific population templates as the mean training-trial activity vector for left and right choices. For each held-out trial, squared Euclidean distances to the two templates were converted to posterior probabilities with a softmax transform. The decision variable was defined as

$$DV_t = \log \frac{P_t(R)}{P_t(L)}.$$

Positive values therefore indicate momentary population evidence for a right choice, negative values indicate evidence for a left choice, and values near zero indicate weak separation between the two decoded choice states.

For Fig. 2e, low-certainty trials were ranked by minimum adjusted P across pre-action bins exceeding 0.1 raw DV units toward the unchosen side, with concordant deviations from the permutation mean. High-certainty selection successively minimized contiguous excursion runs, bin counts and peak amplitude. Choice labels were shuffled within consecutive reward-contingency epochs $\times$ fixed cross-validation folds; templates were refitted without altering neural time series. At each trial-bin location, observed and permuted DV values were

816 standardized using pooled means and SDs. P values used permutation maxima of absolute
817 standardized deviations across all session trials and pre-action bins, with a plus-one
818 correction. This reference assumes label exchangeability within strata under the null.
819 Additional Bonferroni correction covered the 22 sessions searched. Normalization and
820 smoothing served visualization only. For Fig. 2f, unsmoothed DV sign changes outside a ± 0.1
821 dead zone were counted during $-0.8 \leq t < 0$ s relative to Cport retraction. Rates (sign changes
822 per second) were averaged by certainty within each session and region. Linear mixed-effects
823 models included certainty as a fixed effect, rat as a random intercept and session as a nested
824 variance component.

825 *Control analyses for posterior-DV switching.* We performed additional controls to test whether
826 the higher low-certainty switching rate reflected genuine alternation of decoded choice
827 evidence rather than weak or noisy decoder output. First, switching rates were recomputed
828 across multiple DV thresholds defined from the empirical distribution of $|DV|$ within each
829 session and region; the plotted thresholds correspond to the 10th–40th percentiles of pre-
830 action $|DV|$ values (Supplementary Fig. 2a). Second, low- and high-certainty trials were
831 matched by mean pre-action $|DV|$ amplitude before recomputing switching rates, testing
832 whether the effect persisted after equating decoded signal strength (Supplementary Fig. 2b).
833 Third, true-decoder pre-action $|DV|$ amplitude was compared with a null distribution obtained
834 by retraining the posterior decoder after shuffling choice labels 500 times per session
835 (Supplementary Fig. 2c). Fourth, the observed low-minus-high switching-rate difference was
836 compared with a certainty-label-shuffle null in which low/high certainty labels were shuffled
837 within session while preserving decoded DV traces and group sizes (Supplementary Fig. 2d).
838 Control analyses were summarized at the session level and tested with linear mixed-effects
839 models using rat identity as a random effect.

840 *Single-trial outcome-decoding traces.* To track post-outcome population signals, we trained
841 linear SVM classifiers to decode current-trial outcome from mPFC and vIOFC activity aligned
842 to reward-port retraction. Spikes were counted in 50-ms windows stepped every 25 ms across
843 the post-outcome reward-port holding period from -1.0 to 0 s relative to reward-port retraction.
844 Outcome labels were encoded as reward versus no reward. For each neuron, spike counts
845 were z-scored across trials and time bins within session. For each held-out trial and time bin,
846 SVMs were trained on the remaining trials using balanced reward/no-reward training subsets,
847 and the held-out distance to the decision boundary was averaged across folds. Neuron counts
848 were matched between regions by random subsampling, and decoding traces were averaged
849 across 20 subsampling repeats. The signed SVM distance-to-boundary was used as a
850 continuous outcome signal: negative values corresponded to no-reward-associated activity
851 patterns, and positive values corresponded to reward-associated activity patterns. Single-trial
852 traces and condition-mean traces for an example session are shown in Fig. 4c.

853 *Inter-areal temporal-offset analysis of outcome signals.* To quantify the temporal relationship
854 between mPFC and vIOFC outcome signals, we analyzed single-trial SVM distance-to-
855 boundary traces during the post-outcome window. Traces were smoothed with a Gaussian
856 kernel before temporal-offset estimation. For each unrewarded trial, the mPFC and vIOFC
857 traces were mean-centered, and normalized cross-correlation was computed across
858 candidate lags from -0.4 to 0.4 s. A continuous lag estimate was obtained from the center of
859 mass of the positive cross-correlation profile. Negative lag values indicate vIOFC-preceding
860 mPFC dynamics, whereas positive values indicate mPFC-preceding vIOFC dynamics. This
861 analysis estimates temporal precedence between regional population signals and does not

862 establish causal directionality. The procedure is illustrated for an example unrewarded trial in
863 Fig. 4d.

864 For the session-level analysis, we focused on unrewarded trials and grouped them by the
865 certainty of the just-completed choice and by whether the animal stayed or switched on the
866 next trial. High-certainty stay and switch groups were restricted to unrewarded trials after
867 model-concordant choices, defined as choices matching the action assigned higher probability
868 by the RFLR model. Low-certainty stay and switch groups were retained as low-certainty
869 unrewarded-trial comparison groups. Trial-wise lag values were averaged within each session
870 and condition, and planned comparisons were tested with linear mixed-effects models using
871 rat identity as a random effect (Fig. 4e). The schematic summary of this analysis is shown in
872 Fig. 4f.

873 Ensemble Analysis

874 *Population activity preprocessing.* Ensemble analyses were performed separately for the pre-
875 action Cport-hold period and the post-outcome Rport-hold period. For Cport analyses, spike
876 activity was aligned to Cport retraction and analyzed from -0.8 to 0 s. For Rport analyses,
877 activity was aligned to Rport retraction and analyzed from -1.0 to 0 s. Spike trains were
878 converted to 50-ms binned activity, and each neuron's binned activity was z-scored within
879 session. Ensemble detection was performed separately for each rat using stable single units
880 recorded across the sessions included in that analysis. For each rat and epoch, normalized
881 activity matrices from mPFC and vIOFC were stacked and concatenated across sessions to
882 form a neuron $\times$ time-sample population matrix.

883 *PCA-ICA ensemble detection.* Neuronal ensembles were identified with a two-stage PCA-
884 ICA procedure. PCA was first applied to the neuron $\times$ time-sample population matrix to identify
885 dominant modes of correlated population activity. Significant principal components were
886 selected using the Marchenko-Pastur upper edge, $\lambda_{\max} = (1 + \sqrt{N/T})^2$, where N is the number
887 of neurons and T is the number of time samples. Components with eigenvalues above this
888 bound were retained. Independent component analysis was then applied to the retained
889 principal-component subspace using FastICA with unit-variance whitening and
890 max_iter=1000. The number of ICA components was set equal to the number of retained
891 principal components. Each ICA weight vector defined one ensemble. Weight-vector polarity
892 was fixed by flipping the vector when necessary so that the neuron with the largest absolute
893 weight had a positive coefficient.

894 *Ensemble membership and activation.* For visualization and regional-composition analyses,
895 ensemble members were defined as neurons with positive weights exceeding the component
896 mean plus one standard deviation. Inter-areal ensembles were defined as ensembles
897 containing at least one above-threshold mPFC member and at least one above-threshold
898 vIOFC member. Time-resolved ensemble activation was quantified with a quadratic projection,

899 $R_k(t) = (v_k^T z(t))^2 - \sum_i v_{ik}^2 z_i(t)^2$, where v_k is the k th ensemble weight vector and $z(t)$ is the

900 z-scored population activity vector. The diagonal subtraction removes single-neuron power,
901 so the resulting activation reflects coordinated co-activation among neurons contributing to
902 the ensemble.

903 *Cport choice- and certainty-related ensembles.* To identify inter-areal ensembles related to
904 upcoming choice and model-derived choice-state certainty during the Cport-hold period,
905 ensemble activation was tested in sliding 250-ms windows, corresponding to five 50-ms bins.

For each inter-areal ensemble, activity was compared between upcoming left- and right-choice trials using label-permutation tests with a minimum of 10 trials per choice group. The ensemble's preferred choice was defined as the choice associated with higher activation in the selected window. Certainty modulation was then tested in the same window, restricted to trials corresponding to the ensemble's preferred choice. Low- and high-certainty trials were defined within session as the bottom and top tertiles of model-derived choice-state certainty. Ensembles passing both the upcoming-choice and certainty-modulation criteria were used for the four-condition Cport ensemble summaries in Fig. 3a–c.

Corrected screen for time-flexible Cport choice ensembles. For analyses of competing choice-selective ensembles during the Cport-hold period, we used a corrected time-flexible screening procedure. This procedure preserves the possibility that choice-selective ensemble activity can emerge at different times within the Cport epoch while controlling for the sliding-window search. For each inter-areal ensemble, the absolute left-versus-right activation difference was computed across all 250-ms windows, and the maximum absolute effect was used as the observed statistic. A null distribution was generated by shuffling choice labels within session 1,000 times and recomputing the maximum absolute effect across windows for each shuffle. Ensembles with max-statistic-corrected $P < 0.05$ were classified as choice-selective inter-areal ensembles for the competitive-representation analyses in Fig. 3d,e and Supplementary Fig. 3.

Cport ensemble competition metrics. For each corrected choice-selective ensemble, activation traces were z-scored across trials and time bins and smoothed with a Gaussian filter, $\sigma = 1.2$ bins. Ensembles were grouped by preferred choice. On each trial, mean activity of left-preferring and right-preferring ensembles was computed, and activity was then expressed relative to the animal's actual upcoming choice. Chosen-side ensemble dominance was defined as the mean activity of ensembles preferring the chosen action minus the mean activity of ensembles preferring the nonchosen action. A companion binary metric, choice-congruent ensemble activation, was assigned 1 when chosen-side activity exceeded nonchosen-side activity and 0 otherwise. Trial-level metrics were averaged within session for low- and high-certainty trials and compared with linear mixed-effects models using certainty state as a fixed effect and rat identity as a random intercept (Fig. 3e). As an additional validation of ensemble membership, preferred-choice modulation of high-weight ensemble members was compared with that of nonmembers using z-scored single-unit activity in the selected Cport window (Supplementary Fig. 3).

Rport loss-responsive ensembles. Post-outcome ensembles were detected with the same PCA–ICA procedure applied to activity aligned to Rport retraction from -1.0 to 0 s. For visualization and characterization of loss-responsive ensembles, activation was averaged separately for rewarded and unrewarded trials. Candidate loss-responsive inter-areal ensembles were required to contain members in both regions and to be vOFC-dominant, defined as at least half of above-threshold ensemble members belonging to vOFC. Loss reactivity was assessed in three predefined subwindows spanning the Rport epoch: -1.0 to -0.67 s, -0.67 to -0.33 s and -0.33 to 0 s. Within each window, activation on unrewarded and rewarded trials was compared with a two-sided Mann–Whitney U test; the smallest P value across windows was Bonferroni-corrected for the three tested windows. Ensembles with higher activation on unrewarded than rewarded trials after correction were classified as loss-responsive. Representative loss-responsive vOFC-dominant ensembles were ordered by their response timing and plotted as mean $\pm$ s.e.m. activation traces with corresponding neuronal weight distributions in Fig. 4a,b and Supplementary Fig. 4.

References

1. Soltani, A. & Izquierdo, A. Adaptive learning under expected and unexpected uncertainty. *Nat Rev Neurosci* **20**, 635–644 (2019).
2. 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).
3. Uddin, L. Q. Cognitive and behavioural flexibility: neural mechanisms and clinical considerations. *Nat Rev Neurosci* **22**, 167–179 (2021).
4. Vaghi, M. M. *et al.* Specific Frontostriatal Circuits for Impaired Cognitive Flexibility and Goal-Directed Planning in Obsessive-Compulsive Disorder: Evidence From Resting-State Functional Connectivity. *Biological Psychiatry* **81**, 708–717 (2017).
5. Fellows, L. K. & Farah, M. J. Ventromedial frontal cortex mediates affective shifting in humans: evidence from a reversal learning paradigm. *Brain* **126**, 1830–1837 (2003).
6. Klein-Flügge, M. C., Bongioanni, A. & Rushworth, M. F. S. Medial and orbital frontal cortex in decision-making and flexible behavior. *Neuron* **110**, 2743–2770 (2022).
7. Rudebeck, P. H. *et al.* Frontal Cortex Subregions Play Distinct Roles in Choices between Actions and Stimuli. *J. Neurosci.* **28**, 13775–13785 (2008).
8. Wilson, R. C., Takahashi, Y. K., Schoenbaum, G. & Niv, Y. Orbitofrontal Cortex as a Cognitive Map of Task Space. *Neuron* **81**, 267–279 (2014).
9. Findling, C. *et al.* Brain-wide representations of prior information in mouse decision-making. *Nature* **645**, 192–200 (2025).
10. Minxha, J., Adolphs, R., Fusi, S., Mamelak, A. N. & Rutishauser, U. Flexible recruitment of memory-based choice representations by the human medial frontal cortex. *Science* **368**, eaba3313 (2020).
11. Semedo, J. D., Zandvakili, A., Machens, C. K., Yu, B. M. & Kohn, A. Cortical Areas Interact through a Communication Subspace. *Neuron* **102**, 249–259.e4 (2019).
12. Semedo, J. D. *et al.* Feedforward and feedback interactions between visual cortical areas use different population activity patterns. *Nat Commun* **13**, 1099 (2022).
13. Fries, P. Rhythms for Cognition: Communication through Coherence. *Neuron* **88**, 220–235 (2015).
14. Schneider, M. *et al.* A mechanism for inter-areal coherence through communication based on connectivity and oscillatory power. *Neuron* **109**, 4050–4067.e12 (2021).
15. Stoll, F. M. & Rudebeck, P. H. Decision-making shapes dynamic inter-areal communication within macaque ventral frontal cortex. *Current Biology* **34**, 4526–4538.e5 (2024).
16. Beron, C. C., Neufeld, S. Q., Linderman, S. W. & Sabatini, B. L. Mice exhibit stochastic and efficient action switching during probabilistic decision making. *Proceedings of the National Academy of Sciences* **119**, e2113961119 (2022).
17. Kepecs, A., Uchida, N., Zariwala, H. A. & Mainen, Z. F. Neural correlates, computation and behavioural impact of decision confidence. *Nature* **455**, 227–231 (2008).
18. Lak, A. *et al.* Orbitofrontal Cortex Is Required for Optimal Waiting Based on Decision Confidence. *Neuron* **84**, 190–201 (2014).
19. Bradfield, L. A., Dezfouli, A., van Holstein, M., Chieng, B. & Balleine, B. W. Medial Orbitofrontal Cortex Mediates Outcome Retrieval in Partially Observable Task Situations. *Neuron* **88**, 1268–1280 (2015).
20. Tang, H., Bartolo, R. & Averbeck, B. B. Reward-related choices determine information timing and flow across macaque lateral prefrontal cortex. *Nat Commun* **12**, 894 (2021).

- 998 21. Gore, F. *et al.* Orbitofrontal cortex control of striatum leads economic decision-making. *Nat*
999 *Neurosci* **26**, 1566–1574 (2023).
- 1000 22. Sul, J. H., Kim, H., Huh, N., Lee, D. & Jung, M. W. Distinct Roles of Rodent Orbitofrontal and
1001 Medial Prefrontal Cortex in Decision Making. *Neuron* **66**, 449–460 (2010).
- 1002 23. Yasar, T. B. *et al.* Months-long tracking of neuronal ensembles spanning multiple brain areas
1003 with Ultra-Flexible Tentacle Electrodes. *Nat Commun* **15**, 4822 (2024).
- 1004 24. Cools, R., Clark, L., Owen, A. M. & Robbins, T. W. Defining the Neural Mechanisms of
1005 Probabilistic Reversal Learning Using Event-Related Functional Magnetic Resonance
1006 Imaging. *J. Neurosci.* **22**, 4563–4567 (2002).
- 1007 25. van de Ven, G. M., Trouche, S., McNamara, C. G., Allen, K. & Dupret, D. Hippocampal
1008 Offline Reactivation Consolidates Recently Formed Cell Assembly Patterns during Sharp
1009 Wave-Ripples. *Neuron* **92**, 968–974 (2016).
- 1010 26. Umbach, G., Tan, R., Jacobs, J., Pfeiffer, B. E. & Lega, B. Flexibility of functional neuronal
1011 assemblies supports human memory. *Nat Commun* **13**, 6162 (2022).
- 1012 27. Rich, E. L. & Wallis, J. D. Decoding subjective decisions from orbitofrontal cortex. *Nat Neurosci*
1013 **19**, 973–980 (2016).
- 1014 28. Friston, K. *et al.* Active inference and epistemic value. *Cogn Neurosci* **6**, 187–214 (2015).
- 1015 29. Gottlieb, J., Oudeyer, P.-Y., Lopes, M. & Baranes, A. Information-seeking, curiosity, and
1016 attention: computational and neural mechanisms. *Trends in Cognitive Sciences* **17**, 585–593
1017 (2013).
- 1018 30. Cisek, P. Cortical mechanisms of action selection: the affordance competition hypothesis.
1019 *Philos Trans R Soc Lond B Biol Sci* **362**, 1585–1599 (2007).
- 1020 31. Buschman, T. J., Denovellis, E. L., Diogo, C., Bullock, D. & Miller, E. K. Synchronous
1021 Oscillatory Neural Ensembles for Rules in the Prefrontal Cortex. *Neuron* **76**, 838–846 (2012).
- 1022 32. Rassi, E. *et al.* Distinct beta frequencies reflect categorical decisions. *Nat Commun* **14**, 2923
1023 (2023).
- 1024 33. Dias, R., Robbins, T. W. & Roberts, A. C. Dissociation in prefrontal cortex of affective and
1025 attentional shifts. *Nature* **380**, 69–72 (1996).
- 1026 34. Öngür, D. & Price, J. L. The Organization of Networks within the Orbital and Medial
1027 Prefrontal Cortex of Rats, Monkeys and Humans. *Cereb Cortex* **10**, 206–219 (2000).
- 1028 35. Euston, D. R., Gruber, A. J. & McNaughton, B. L. The Role of Medial Prefrontal Cortex in
1029 Memory and Decision Making. *Neuron* **76**, 1057–1070 (2012).
- 1030 36. Bari, B. A. *et al.* Stable Representations of Decision Variables for Flexible Behavior. *Neuron*
1031 **103**, 922-933.e7 (2019).
- 1032 37. Romero-Sosa, J. L., Yeghikian, A., Wikenheiser, A. M., Blair, H. T. & Izquierdo, A. Neural
1033 coding of choice and outcome are modulated by uncertainty in orbitofrontal but not secondary
1034 motor cortex. *Nat Commun* **16**, 8931 (2025).
- 1035 38. Masset, P., Ott, T., Lak, A., Hirokawa, J. & Kepecs, A. Behavior- and Modality-General
1036 Representation of Confidence in Orbitofrontal Cortex. *Cell* **182**, 112-126.e18 (2020).
- 1037 39. Buzsáki, G. Neural Syntax: Cell Assemblies, Synapsembles, and Readers. *Neuron* **68**, 362–
1038 385 (2010).
- 1039 40. Yuste, R., Cossart, R. & Yaksi, E. Neuronal ensembles: Building blocks of neural circuits.
1040 *Neuron* **112**, 875–892 (2024).
- 1041 41. Carrillo-Reid, L., Yang, W., Bando, Y., Peterka, D. S. & Yuste, R. Imprinting and recalling
1042 cortical ensembles. *Science* **353**, 691–694 (2016).
- 1043 42. Fujisawa, S., Amarasingham, A., Harrison, M. T. & Buzsáki, G. Behavior-dependent short-
1044 term assembly dynamics in the medial prefrontal cortex. *Nat Neurosci* **11**, 823–833 (2008).

- 1045 43. Benchenane, K. *et al.* Coherent Theta Oscillations and Reorganization of Spike Timing in the
1046 Hippocampal- Prefrontal Network upon Learning. *Neuron* **66**, 921–936 (2010).
- 1047 44. Jadhav, S. P., Rothschild, G., Roumis, D. K. & Frank, L. M. Coordinated Excitation and
1048 Inhibition of Prefrontal Ensembles during Awake Hippocampal Sharp-Wave Ripple Events.
1049 *Neuron* **90**, 113–127 (2016).
- 1050 45. Spellman, T. *et al.* Hippocampal–prefrontal input supports spatial encoding in working memory.
1051 *Nature* **522**, 309–314 (2015).
- 1052 46. Young, R. A., Shin, J. D., Guo, Z. & Jadhav, S. P. Hippocampal–Prefrontal Communication
1053 Subspaces Align with Behavioral and Network Patterns in a Spatial Memory Task. *eNeuro* **12**,
1054 (2025).
- 1055 47. Li, J., Tang, L., Wei, X., Chen, Y. & Xu, H. Flexible goal learning involves coordinated
1056 population activity in dCA1 and medial orbitofrontal cortex. *PLOS Biology* **24**, e3003824
1057 (2026).
- 1058 48. Schiereck, S. S. *et al.* The orbitofrontal cortex updates beliefs for state inference. *Neuron* **114**,
1059 507–520.e8 (2026).
- 1060 49. Constantinople, C. M. *et al.* Lateral orbitofrontal cortex promotes trial-by-trial learning of risky,
1061 but not spatial, biases. *eLife* **8**, e49744 (2019).
- 1062 50. Gardner, M. P. H. *et al.* Processing in Lateral Orbitofrontal Cortex Is Required to Estimate
1063 Subjective Preference during Initial, but Not Established, Economic Choice. *Neuron* **108**, 526–
1064 537.e4 (2020).
- 1065 51. Banerjee, A. *et al.* Value-guided remapping of sensory cortex by lateral orbitofrontal cortex.
1066 *Nature* **585**, 245–250 (2020).
- 1067 52. Lichtenberg, N. T. *et al.* Basolateral Amygdala to Orbitofrontal Cortex Projections Enable Cue-
1068 Triggered Reward Expectations. *J Neurosci* **37**, 8374–8384 (2017).
- 1069 53. Jenni, N. L., Bercovici, D. A. & Floresco, S. B. Medial Orbitofrontal, Prefrontal, and
1070 Amygdalar Circuits Support Dissociable Component Processes of Risk/Reward Decision-
1071 Making. *J. Neurosci.* **45**, (2025).
- 1072 54. Siegel, M., Buschman, T. J. & Miller, E. K. Cortical information flow during flexible
1073 sensorimotor decisions. *Science* **348**, 1352–1355 (2015).
- 1074 55. Friston, K. The free-energy principle: a unified brain theory? *Nat Rev Neurosci* **11**, 127–138
1075 (2010).
- 1076 56. Kepecs, A. & Mainen, Z. F. A computational framework for the study of confidence in humans
1077 and animals. *Philos Trans R Soc Lond B Biol Sci* **367**, 1322–1337 (2012).
- 1078 57. Nassar, M. R. & Gold, J. I. A Healthy Fear of the Unknown: Perspectives on the
1079 Interpretation of Parameter Fits from Computational Models in Neuroscience. *PLOS*
1080 *Computational Biology* **9**, e1003015 (2013).
- 1081 58. Schoonover, C. E., Ohashi, S. N., Axel, R. & Fink, A. J. P. Representational drift in primary
1082 olfactory cortex. *Nature* **594**, 541–546 (2021).
- 1083 59. Jun, J. J. *et al.* Real-time spike sorting platform for high-density extracellular probes with
1084 ground-truth validation and drift correction. 101030 Preprint at <https://doi.org/10.1101/101030>
1085 (2017).
- 1086 60. Jiang, X., Ryu, S. I., Shenoy, K. V. & Kao, J. C. Single Neuron Firing Rate Statistics in Motor
1087 Cortex During Execution and Observation of Movement. in *2018 40th Annual International*
1088 *Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* 981–986 (2018).

Acknowledgements

This project has been funded by the Swiss Federal Institute of Technology (ETH) Zurich (internal funds, M.F.Y.), the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 818179, M.F.Y.), the Swiss National Science Foundation (SNSF; Sinergia Project CRSII5_198739, Advanced Grant TMAG-3_226025, and SNSF/Innosuisse BRIDGE Discovery 40B2-0_211760, M.F.Y.), an IEEE Brain Student International Travel Grant, and a Swiss Society for Neuroscience Travel Fellowship. We thank Abbas Khani for critically inspecting the data-preparation and curation workflow and diagnosing problems before analysis. We thank Houman Javaheri for his exploratory work on spike-sorting approaches for tracking neurons across days. We thank Thomas Gent for helping to refine the surgical technique for obtaining long-term stable recordings in rats. We thank Theofanis Karayannis and Benjamin Grewe for their valuable feedback and support throughout the project. We also thank Abhishek Banerjee for valuable feedback that helped improve the manuscript.

Author contributions

Z.F. developed the analysis software and implemented the analysis pipelines, produced the final versions of the figures, and revised the manuscript. B.Wu conceptualized and designed the study, led and supervised the project, performed the animal experiments, collected the electrophysiological recordings, curated the data, and wrote the original draft of the manuscript and the initial versions of the figures. Z.F. and B.Wu jointly designed the analysis approach, performed the behavioral modeling and neural-data analyses, interpreted the results, and reviewed and edited the manuscript. B.Wu and W.B. performed the electrode implantation surgeries in rats. T.B.Y. designed and implemented the ultra-flexible tentacle electrodes used for longitudinal, simultaneous recordings from the medial prefrontal and ventrolateral orbitofrontal cortices. E.Ö. contributed to the design and fabrication of the ultra-flexible tentacle electrodes used in this study. P.G. introduced the temporal-activity-correlation-based PCA–ICA pipeline for neuronal ensemble detection and devised its extension to multi-area data for identifying stable inter-areal co-activation patterns. P.G. also advised on other analytical methods for inter-areal data. W.B. and M.F.Y. contributed to the study design and data interpretation and advised on manuscript writing. M.F.Y. supervised the study as its principal investigator and corresponding senior author. All authors reviewed and approved the final manuscript.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [OFCmPFCNatureNeurosciencesupplementary.pdf](#)