

Orbitofrontal Cortex Signals Expected Outcomes with Predictive Codes When Stable Contingencies Promote the Integration of Reward History

 Justin S. Riceberg and Matthew L. Shapiro

Department of Neuroscience, Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029

Memory can inform goal-directed behavior by linking current opportunities to past outcomes. The orbitofrontal cortex (OFC) may guide value-based responses by integrating the history of stimulus–reward associations into expected outcomes, representations of predicted hedonic value and quality. Alternatively, the OFC may rapidly compute flexible “online” reward predictions by associating stimuli with the latest outcome. OFC neurons develop predictive codes when rats learn to associate arbitrary stimuli with outcomes, but the extent to which predictive coding depends on most recent events and the integrated history of rewards is unclear. To investigate how reward history modulates OFC activity, we recorded OFC ensembles as rats performed spatial discriminations that differed only in the number of rewarded trials between goal reversals. The firing rate of single OFC neurons distinguished identical behaviors guided by different goals. When >20 rewarded trials separated goal switches, OFC ensembles developed stable and anticorrelated population vectors that predicted overall choice accuracy and the goal selected in single trials. When <10 rewarded trials separated goal switches, OFC population vectors decorrelated rapidly after each switch, but did not develop anticorrelated firing patterns or predict choice accuracy. The results show that, whereas OFC signals respond rapidly to contingency changes, they predict choices only when reward history is relatively stable, suggesting that consecutive rewarded episodes are needed for OFC computations that integrate reward history into expected outcomes.

Key words: decision making; neurophysiology; orbitofrontal cortex; reversal learning; reward prediction; single units

Significance Statement

Adapting to changing contingencies and making decisions engages the orbitofrontal cortex (OFC). Previous work shows that OFC function can either improve or impair learning depending on reward stability, suggesting that OFC guides behavior optimally when contingencies apply consistently. The mechanisms that link reward history to OFC computations remain obscure. Here, we examined OFC unit activity as rodents performed tasks controlled by contingencies that varied reward history. When contingencies were stable, OFC neurons signaled past, present, and pending events; when contingencies were unstable, past and present coding persisted, but predictive coding diminished. The results suggest that OFC mechanisms require stable contingencies across consecutive episodes to integrate reward history, represent predicted outcomes, and inform goal-directed choices.

Introduction

Memory informs choices by linking current opportunities to the outcome of past actions. When contingencies change, animals

learn to repeat rewarded responses and avoid other ones. Because contingencies change at different rates, optimal responses require tracking the reliability of outcomes. Across species, the orbitofrontal cortex (OFC) supports adaptive choices based on reward, but its specific function is debated. The OFC may track contingency changes rapidly and guide choices by remapping stimulus–reward associations after every trial (Rolls, 2004). Alternatively, the OFC may integrate reward history, accumulating evidence over several trials, and guide choices by computing expected outcomes (Schoenbaum et al., 2009).

OFC dysfunction impairs performance when stable contingencies change and reward expectancies are violated (Baxter et

Received Sept. 20, 2016; revised Dec. 8, 2016; accepted Dec. 29, 2016.

Author contributions: J.S.R. and M.L.S. designed research; J.S.R. performed research; J.S.R. and M.L.S. analyzed data; J.S.R. and M.L.S. wrote the paper.

This work was supported by the National Institutes of Health (Grants MH065658, MH065658, and MH052090), the Icahn School of Medicine at Mount Sinai, and the NVIDIA Corporation. We thank Kevin Guise, Pierre Enel, Pablo Martin, and Vincent Luo for helpful comments on the manuscript and Anna Balk and Maojuan Zhang for technical help.

The authors declare no competing financial interests.

Correspondence should be addressed to Justin S. Riceberg, Department of Neuroscience, Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, 1 Gustave Levy Place, Box 1065, New York, NY 10029. E-mail: justin.riceberg@mssm.edu.

DOI:10.1523/JNEUROSCI.2951-16.2016

Copyright © 2017 the authors 0270-6474/17/372010-12\$15.00/0

al., 2000, Pickens et al., 2003). OFC function may be less useful (Stalnaker et al., 2015) or detrimental (Riceberg and Shapiro, 2012) to learning when contingencies change often. OFC unit activity predicts choices when rats and monkeys learn to associate stimuli with rewards, but the effects of different rates of contingency changes on such activity are not well understood. OFC firing correlates are typically analyzed during different phases of operant tasks: stimulus presentation, delay, reward availability, and consumption. In rats, different populations of OFC units become active in each of these phases (van Wingerden et al., 2010a;2010b). Neuronal responses to stimulus presentation that distinguish pending rewards are considered representations of expected outcome. As rats learn odor-flavor associations, OFC single units gain predictive signals, fire in distinct phases of theta and gamma oscillations, and the predictive codes vanish immediately after contingencies change (van Wingerden et al., 2010a, 2010b, 2014). OFC neurons in monkeys taught to use visual fixation cues to direct saccades developed predictive activity over repeated trials that diminished when contingencies destabilized and disappeared when they were violated (Kobayashi et al., 2010). OFC circuits seem to respond to consistent reward associations by accumulating evidence, perhaps by integrating the history of episodes in a cognitive map of task space, a neuronal representation of predicted outcomes that lets inferred values inform decisions (Wilson et al., 2014).

Rats associate places with hidden food and safety and adapt rapidly to changing spatial contingencies, yet the OFC's contribution to spatial learning has not been investigated extensively. To investigate how reward history affects OFC coding, we recorded OFC ensembles as rats performed identical spatial discriminations and reversals in plus-maze tasks that differed only in reversal frequency. In the plus maze, a rat is placed in a north or south start arm and can obtain hidden food by approaching the choice point and selecting the east or west goal arm. Each start arm is a discriminative stimulus and selecting a goal arm is a discriminative response. If the OFC maps stimulus–reward associations, then OFC neurons should respond rapidly to reward changes and be unaffected by contingency stability. Stimulus–reward remapping should be most prominent in the goal arm, where reward is obtained and contiguity is strongest. Alternatively, if the OFC integrates reward history, then OFC units should respond slowly over successive episodes and stable contingencies. OFC activity can inform choices only before the rat exits the choice point, for example, in the start arm. If outcome expectancies signaled by OFC neurons inform choices, then OFC activity in the start arm should predict goal selection and depend upon stable contingencies. Although OFC activity changed rapidly after contingency changes, OFC ensembles predicted choices only when contingencies were stable, suggesting that OFC circuits compute expected outcomes by integrating reward across successive trials.

Materials and Methods

Experimental design

We compared spatial learning in a plus maze using tasks with different reversal frequencies (Fig. 1A) (Riceberg and Shapiro, 2012). A “low-frequency reversal” (LFR) task was intended to promote stable outcome expectancies by maintaining the same reward contingency and required each rat to perform 17/20 consecutive trials correctly in a block of >20 trials before switching reward contingencies. A “high-frequency reversal” (HFR) task was intended to discourage the formation of stable outcome expectancies by switching reward contingencies frequently and required each rat to perform five of seven consecutive trials correctly in a block of <12 trials. The two tasks thereby favored different response

tendencies, with LFR encouraging the integration of reward history into outcome expectancies and the HFR encouraging the formation of more immediate reward associations. The rats were acclimated to the maze, pretrained, implanted surgically with electrodes, retrained, and OFC neurons were recorded as they performed the LFR and HFR tasks (Fig. 1A). Across all rats, 14 sessions of each task were included for analysis.

Subjects

Adult male Long–Evans rats weighed 275–300 g at the start of experiments and were housed individually in a colony room with a 12 h light/dark cycle. The rats were acclimated to the colony for a week and then food restricted to no <85% of their *ad libitum* body weight and maintained on a restricted diet for the duration of the experiment. Four rats were implanted with tetrode bundles targeting neurons in the ventrolateral OFC (Fig. 1A). All procedures were performed in accordance with Institutional Animal Care and Use Committee guidelines and those established by the National Institutes of Health.

Maze

A plus-shaped maze made of wood was painted gray with 4 arms (59 cm long, 6.5 cm wide, with edges 2 cm high) that met in the center at 90° angles. Food cups at the end of each arm were recessed 0.5 cm and an inaccessible piece of food was kept below a mesh screen. Each of two opposing arms was designated as north and south start arms; the two orthogonal arms were designated east and west goal arms. A rectangular waiting platform made of white-painted wood (30 × 35 cm) stood beside the maze in a room with several visual cues on the walls. The platform and the maze were elevated 81 cm above the floor.

Surgery

Each rat was anesthetized with continuous flow isoflurane, mounted in a stereotaxic frame, and given a preoperative dose of buprenorphine (0.3 mg/kg) intraperitoneally. Body temperature was monitored throughout the surgery. The scalp was shaved, anesthetized (0.5 ml lidocaine with epinephrine, subcutaneous), incised at the midline, and the skin and pericranium were retracted. Seven holes were drilled around the perimeter of the skull for ground screws ($n = 4$) and reference screws ($n = 3$). The skull and screws were coated with a layer of Metabond and Panavia. Craniotomies were made above the right OFC (3.8 mm anterior, 2.2 mm lateral to bregma). Dura was removed, the hyperdrive was lowered so that the electrode guide tubes met the cortical surface, and the hyperdrive was fixed to the skull with dental acrylic. Ground and reference wires were pinned and cemented into an electrode interface board (EIB) with cable connectors for recording. Tetrodes were lowered 2.2–2.5 mm toward the OFC at the end of surgery. Rats were administered sterile saline (0.5 ml/h, s.c.) and given a single postsurgical dose of buprenorphine and daily MetaCam as needed during a >7 d recovery period.

Histology

At the end of testing, rats were deeply anesthetized with isoflurane and pentobarbital (100 mg/ml, i.p.). One wire in each eligible tetrode was pulsed with 300 μ A DC to label the tetrode tips via a small lesion. Rats were perfused transcardially with PBS and then 10% formalin. Brains were removed, cryoprotected in 15% followed by 30% sucrose solution, sectioned at 40 μ m, and stained with formol–thionin to label cell bodies blue and fiber tracts red.

Behavior testing

Pretraining. The rats were handled for 5 d and acclimated to the testing room and the maze, where they found scattered chocolate sprinkles. Inaccessible food rewards were distributed throughout the maze to minimize the use of odor cues to guide foraging. The rats were then trained in a spatial win–stay task. Each trial began when the experimenter placed food in the designated goal arm, picked up a rat from the waiting platform, and placed it at the end of a pseudorandomly selected start arm facing away from the maze center. The experimenter used similar movements while placing food in the food cups to avoid cueing its location to the rat. The rats were trained initially to find chocolate sprinkles at the end of the west arm. Self-correction was permitted during this stage. After the rat consumed the reward, the experimenter picked up the rat

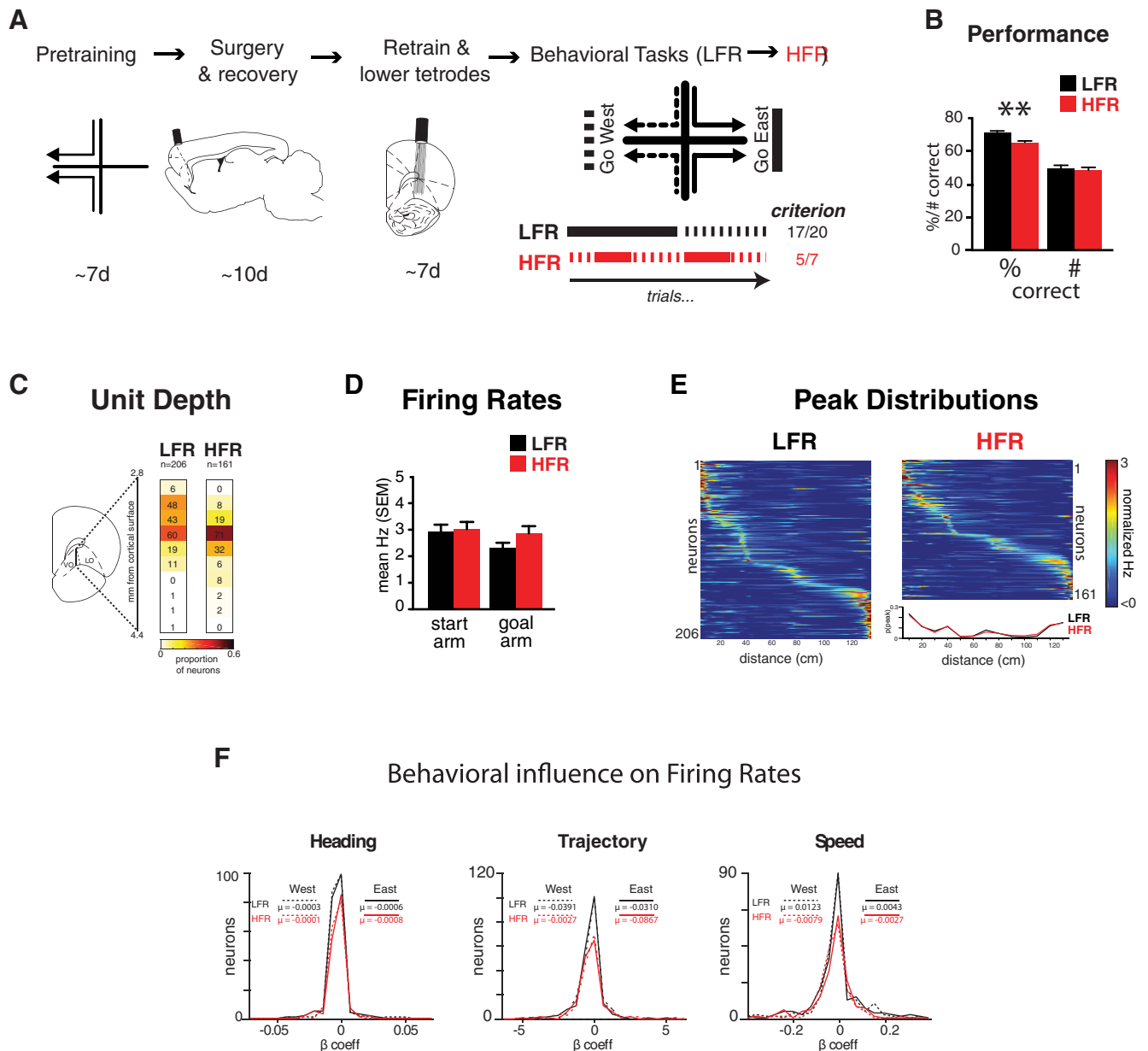


Figure 1. Experimental design, behavior, and firing correlates. **A**, Rats were trained to find food in one goal arm in a plus maze and then implanted with tetrodes targeting ventrolateral OFC. Schematics depict sagittal and coronal brain sections 2.4 mm lateral and 3.7 mm anterior from bregma, respectively. After recovery from surgery, each rat was retrained to find reward in east or west goal arms in blocks of trials: “go west” (dotted black line) or “go east” (solid black line). Below, lines show correct contingencies during the first blocks of typical low and HFR sessions: LFR (black) and HFR (red). **B**, Rats performed the LFR task with 72% and the HFR task with 66% accuracy: the total number of correct trials performed during each task was similar (700 LFR, 688 HFR: ~50 trials/session for each task). **C**, The numbers of neurons recorded during LFR and HFR and their estimated dorsoventral coordinates are shown. **D**, Firing rates across tasks. Single-unit activity was recorded as rats learned LFR and HFR tasks. Mean firing rates in the start and goal arms were equivalent in LFR (black bars) and HFR (red bars). **E**, Spatial distributions of normalized OFC firing rates. Each row shows the mean firing rate of one unit from start to goal in rows sorted by peak firing rate locations. The spatial distributions of firing rate peaks were indistinguishable in LFR and HFR (inset lower right). **F**, Behavioral variation and firing rates. General linear models quantified the effects of heading angle, spatial trajectories, and movement speed on OFC activity. The variance in firing rate predicted by each behavioral variable (β coefficients, horizontal axis) is plotted for every neuron (vertical axis) in LFR (black) and HFR (red) trials. Solid lines show westbound journeys, dotted lines eastbound journeys, and means (μ) are inset. None of the behavioral variables influenced firing rates in either task.

and placed it on the waiting platform for 5–10 s between trials. After a rat made eight consecutive correct “go west” responses, it was assigned to surgery (Fig. 1A).

Postsurgery acclimation and tetrode lowering. After the rat recovered from surgery, it was again acclimated to the maze, retrained in the spatial discrimination task, and the tetrodes were advanced toward their target in 60–180 mm steps once or twice daily (Fig. 1A). On testing days, tetrodes were moved in 20–30 mm increments until stable single-unit activity was detected. Tetrodes were left untouched for at least 1 h before behavioral testing. Final tetrode recording positions are depicted in Figure 1C.

Reversal tasks. As in pretraining, the experimenter put food in the west goal arm: each trial started when the rat was placed in a start arm and ended when the rat was put back on the waiting platform after it either consumed the food or reached the end of the unrewarded arm. A choice was counted when the rat put all four paws into a goal arm. Each rat was given 5–8 d of discrimination training and tetrodes were lowered concurrently toward the OFC. Unit recording began after a rat met the 17/20 correct trial criterion in the “go west” discrimination and stable units were detected. The same contingency held during the first recording sessions and initial block of LFR testing, which was divided into blocks of trials. In each block, the same goal arm was rewarded until the rat met the

criterion performance of 17/20 correct trials (40–97 total trials). The opposite goal arm was rewarded in the next block of trials so that each daily testing session included two to four blocks and one to three reversals. Because our previous work found that the order of HFR and LFR training did not alter the effect of OFC lesions (Riceberg and Shapiro, 2012), the rats were tested on the HFR task after LFR testing. The tasks were procedurally identical except the criterion for contingency change was five of seven consecutive correct trials in HFR (52–93 total trials) (Fig. 1A). Rats completed three to seven reversals per HFR session. Self-correction was not permitted in either task.

Recording and unit discrimination

Recording apparatus. Hyperdrives held between 12 and 24 tetrodes (four twisted nickel-chromium wires, 12.7 μm diameter; RO-800, Kanthal Precision Technology), each loaded into an independently movable microdrive. Each tetrode wire was connected to the EIB and gold plated to an impedance $<0.2\text{ M}\Omega$. The interface board was connected via three 36 channel head stages to a Digital Cheetah data acquisition system to record tetrode signals. Light-emitting diodes were mounted on the head stage and tracked by an overhead video camera (640 \times 480 pixels, 30 Hz sampling rate; Neuralynx). Tetrode signals were digitized and amplified (200–3000 \times) independently for each channel to maximize waveform discrimination. Single units (threshold $>100\text{ mV}$) were sampled for 1 MS at 32,000 Hz and filtered between 600 and 6000 Hz. LFPs were recorded continuously from the same wires sampled at 2000 Hz and filtered between 1 and 512 Hz.

Recording epochs. OFC activity was recorded on each trial from the time the animal was placed on the start arm until it reached the food cup at the end of the goal arm. Recording was paused $\sim 2\text{ s}$ after reward (both reward receipt and omission) and then resumed when the rat was on the wait platform between trials (intertrial interval = $9 \pm 0.16\text{ s}$).

Spike clustering. Spikes were described by waveform parameters; for example, the spike amplitude and duration recorded from each tetrode wire and discriminated as reliable clusters in a high-dimensional parameter space using an eight-step, semiautomated procedure as follows. (1) The initial parameters included the digitized waveform (32 V samples recorded in 1 ms) from each tetrode wire and its Morlet wavelet transform between 600 and 9000 Hz to quantify time-varying amplitude and frequency. (2) Principal components analysis (PCA) of every sample of each tetrode's parameterized waveform recorded in the session identified the three largest independent sources of variance and these, together with any other dimensions with multimodal distributions, defined the parameter space used for cluster identification. (3) Clusters of points in the PCA space were identified initially using KlustaKwik (<http://klustakwik.sourceforge.net/>). (4) Waveforms identified by parameter clusters were inspected visually to eliminate artifacts (e.g., chewing). (5) The remaining parameterized waveforms were clustered again using KlustaKwik. (6) The output was edited using MATLAB routines that displayed 3D projections of the spike waveforms within the remaining clusters. (7) The remaining clusters were refined using auto- and cross-correlations to determine that spike timing was consistent with well discriminated single units. (8) Cluster discrimination was quantified by the Mahalanobis distance.

Unit stability. The stability of individual units was evaluated by measuring drift in mean spike amplitude and firing rate for each tetrode wire in 20 equally distributed time bins spanning each recording session. All units with time-correlated firing rates or waveform amplitudes on any wire were excluded from further analysis (Spearman rank correlation, uncorrected $p < 0.05$). The exclusion criteria were conservative and rejected units that seemed stable by visual inspection, removing 137/504 ($\sim 28\%$) of the recorded OFC units, leaving 367 stable waveforms for analysis.

Identification of cell types. The current analyses targeted OFC pyramidal cells. To distinguish interneurons from pyramidal cells, three neurophysiological parameters were calculated: peak-to-trough duration, peak-to-trough asymmetry, and firing rate (Wilson et al., 1994). The waveform from the tetrode channel with the highest peak amplitude was used to calculate peak to trough time, asymmetry score (peak/trough amplitude), and mean firing rate. Units were clustered into two groups,

one with broader, asymmetric spikes and low mean firing rates and the other with narrower, more symmetrical waveforms and higher firing rates. The first group identified putative pyramidal neurons that were analyzed, whereas the second group identified putative interneurons that were excluded.

Movement variables

Most of the analyses (described next) compared firing patterns in goal-directed journeys as rats followed identical paths either in the same start arm on the way to different goals or in the same goal arm on the way from different starts. Spatial trajectories were constrained physically by the size of the rat ($\sim 4\text{ cm}$ wide) and the narrow maze arms (6.5 cm wide), minimizing potential differences spatial behavior. To quantify the extent to which horizontal head position and movement affected unit activity, we analyzed the variance in movement speed, location occupancy, and heading angle within the spatial ($\sim 0.4\text{ cm}/\text{camera pixel}$) and temporal (30 Hz) sampling resolution of the recording system.

Speed. For each trial, velocities were calculated on linearized position data (see "Trajectory" section), smoothed, and interpolated to successive 1 mm bins. Speed means and SDs were computed for each bin across all trials within sessions. Unit analyses excluded all epochs with speed $<1\text{ cm/s}$, those ± 2.5 SDs from the mean, and all trials in which the rat reversed direction or exhibited erratic behavior (e.g., stopped moving for $>4\text{ s}$).

Trajectory. To quantify variability in spatial trajectories, we compared the distribution of positions occupied by a rat to an idealized path defined by a high-order polynomial curve fit to all trials in a session. Trajectory variance was calculated as the "residual distance" between each tracked point in every trial's actual path and the corresponding points in the idealized path. Residual distances were binned in the start and goal arms of the maze for all trials and each task.

Heading angles. Heading angles were calculated by the arctangent of successive x - y tracking positions for each trial: the angular concentration parameter, κ , quantified the variability of these angles. A generalized linear model quantified the variance in firing rate associated with each of these behavior measures.

Single-unit selectivity

To quantify how the activity of individual OFC neurons varied with task demands, we calculated speed-filtered firing rates of each single unit in every maze region (i.e., start arm, goal arm, reward site, full trial) on all trials and compared the mean firing rate of each unit in different memory epochs and contingencies. Unit activity was sorted into one of four categories by the start and goal of each trial: north or south starts and east- or west-bound journeys. Predictive, retrospective, reward site, and outcome selectivity was assessed by firing rate differences across corresponding journeys: Predictive coding, the extent to which unit activity in a start arm distinguished pending choices to the east or west goal as the rat approached the choice point, was quantified as the difference in firing rates between east- and west-bound journeys, calculated separately for each start arm (e.g., in the north arm during east and west bound journeys). Retrospective coding measured firing rate differences in each goal arm (e.g., in the west goal arm during journeys that started in the north vs south arms). Reward site coding compared firing rate differences in the east and west goal sites and outcome selectivity compared firing rate differences between correct and incorrect trials in identical journeys (e.g., rewarded versus nonrewarded north-to-east journeys). For each unit, the actual difference in mean firing rate between conditions was compared with a null distribution established by permutation tests that shuffled trial identity 5000 \times and by t tests. A significant firing rate difference identified predictive, retrospective, reward-site, and outcome selectivity, and Cohen's d measured the effect size of firing rate differences. A unit was defined operationally as selective if the mean firing rate difference was $>95\%$ of the permuted differences ($p < 0.05$). The proportions of selective units in the two tasks were compared using χ^2 tests. To be included in any analysis, each unit had to fire ≥ 25 spikes during $\geq 20\%$ of trials in a given condition. Firing rates were based on spiking in identical maze regions sampled for similar intervals in both tasks (mean sample time in ms, reward site: HFR = 1050, LFR = 1050;

start arm: HFR = 500, LFR = 550, $t = 1.23$, $p = 0.022$; goal arm: HFR = 820, LFR = 1010 $t = 2.82$, $p < 0.01$).

Population vector correlations

To assess the effects of contingency changes on OFC ensemble activity, we compared the firing rates of simultaneously recorded units as rats move through the start arm and stereotyped behavior was “clamped” before the choice point and goal selection. We reasoned that if OFC neurons signaled pending goals or outcomes, then ensemble activity should be similar from one trial to another as rats follow an established contingency, change when a new contingency is imposed, and stabilize as rats perform to criterion.

Trials for all sessions and rats were grouped by performance according to learning stage. LFR learning curves were generated using a hidden Markov model that estimated the probability of a correct response for each trial across the block of trials defined by a specific contingency (Smith et al., 2004). The Smith algorithm considers the entire session performance to determine the probability correct on a given trial and generates confidence intervals (CIs). For a two-choice task such as the current one, the rat is considered to perform better than chance when the lower confidence bound exceeds 50% probability correct.

During the first block of LFR, rats performed at near ceiling levels and trials were grouped into three evenly spaced sub-blocks (sub-blocks 1, 2, and 3; see Fig. 4A). During the reversal blocks, trials were grouped into quartiles based on the lower CI of the Smith probability: early ≤ 0.2 , middle = 0.2–0.4, late = 0.4–0.6, and stable >0.6 . The HFR task is not well suited to the Smith algorithm because the quickly changing contingencies interfere with measures of stable performance. We therefore grouped the trials within each block into sub-blocks divided by a median split.

Transitions between activity states. We quantified ensemble activity using population vectors (PVs) and compared activity dynamics across the learning curve using PV correlations (PVRs) (Rich et al., 2009). PVs were defined as the mean z-scored firing rates of neurons active on a start arm in $>20\%$ of trials for each ensemble of units recorded simultaneously in every sub-block of each session. To compare similar numbers of sessions in each treatment and trials to each goal, we analyzed PVs from the first two blocks of 10 LFR sessions and the first four blocks of 11 HFR sessions. PVs recorded during performance of the first contingency defined a “seed” used to compute PVRs across contingency changes. In LFR, the seed PV included all trials performed during stable performance of the first contingency; in HFR, the seed PV included all trials in the first contingency block. To quantify the decay of established representations, PVRs compared the initial seed to PVs in subsequent sub-blocks. To quantify the emergence of new activity patterns, new seeds were defined by PVs recorded at the end of each contingency, during late learning and stable performance in LFR, and the entire final contingency block in HFR.

Because the number of contingency changes differed in the LFR and HFR, we also measured PVRs defined by different sub-block and seed definitions. To test the effect of sampling procedures on PVRs, we exchanged the methods used to group LFR and HFR trials. One test grouped all trials after the Smith learning trial (lower CI > 0.5) to define the seed for LFR sessions; another divided LFR trials into split quarters rather than by the Smith algorithm. Complementary tests applied to HFR sessions, for example, defining the initial seed by the trials needed to reach the same percentage correct (five of seven correct) as in HFR and defining one HFR seed for each of the four contingency blocks. To assess the effect of individual trials and trial numbers within sub-blocks, we calculated PVs in 50 randomly drawn sets of five trials for sub-blocks containing more than five trials and used the mean of the trial-randomized PVs in additional PVRs.

Linear classifiers

To test the effect of contingency stability on OFC goal decoding, we used a linear classifier to quantify how well neuronal activity in the start arm predicted goal choices in single trials as rats attained five of seven correct in each goal (Rich and Wallis, 2016). A moving window of seven trials was convolved with the performance vector for the relevant block. The

first five correct trials within that window were selected as the first block of five of seven trials. The same procedure was used after the contingency change to select the first five correct of 10 trials for block 2. We compared decoding accuracy in both tasks and in both individual rats and all rats combined. For each task, the firing rates of each neuron in every recording session were assembled into pseudo-PVs (PPVs) and the 10 trial \times N neuron arrays were used to predict choices using leave-one-out cross-validation. Because many neurons (20–59) predicted the outcome of relatively few trials (10), we used lasso regressions with leave-one-out cross-validation (Hastie et al., 2009). This method removes the data of the trial to-be-classified and calculates a penalty term λ through a leave-one-out cross-validation using the remaining nine trials. The λ -optimized model then classifies the activity on the omitted trial and calculates an accuracy score (0–1) using a logistic function. The overall accuracy of each model is measured by the sum of squares error calculated for the 10 trials. To compare actual to chance predictions, goal labels were shuffled and the same lasso regression method was used to classify the permuted trials. This procedure was repeated 700 times for each individual rat and 200 times for the combined dataset. The probability of obtaining the actual decoding accuracy relative to chance (p -value) was calculated by comparing the actual sum-square-error to the distribution of sum-squared errors in the shuffled data. PPVs from individual rats included neurons that fired in at least three of five correct trials in one or both goal blocks (neurons/rat 1, 2, 3: LFR = 26, 30, 22; HFR = 25, 19, 22, respectively). The combined PPVs included all neurons in every rat that fired in at least four of five correct trials in one or both goal blocks (LFR = 58, HFR = 47 neurons). The different inclusion criteria meant that the total number of neurons differed in the individual and combined analyses and both procedures give similar results (see Fig. 4B).

Results

Variations in spatial movement did not predict firing rates

Because running speed, spatial location, and heading can modulate neural activity, we assessed the extent to which these behavioral variables predicted firing rate dynamics within and between tasks. We calculated variations in firing rate, speed, location, and heading in the start arms for each trial and used GLM to quantify the contribution of these variables to the firing rate of every neuron. None of the behavioral variables predicted firing rate, none of the distribution means differed from zero (range of statistical values: LFR, $t_{(206)} = 0.57$ – 1.58 , $p = 0.12$ – 0.27 ; HFR, $t_{(161)} = 0.85$ – 1.1 , $p = 0.31$ – 0.38), and none of the measures distinguished firing rates in either of the two tasks [LFR vs HFR: heading angle (κ), Kolmogorov–Smirnov statistic (ks-stat) = 0.13, $p = 0.09$; trajectory, ks-stat = 0.09, $p = 0.45$; running speed, ks-stat = 0.11, $p = 0.23$; Fig. 1F] or the trial’s destination (east vs west: LFR, $k = 0.04$ – 0.09 , $p = 0.4$ – 0.97 ; HFR, $k = 0.08$ – 0.11 , $p = 0.31$ – 0.65 ; Fig. 1F).

Reversal frequency altered learning but not firing rate

All rats were trained initially to enter one spatial goal from each of the two start arms (e.g., north-to-west and south-to-west). On subsequent testing days, the reward contingencies changed without warning so the rat had to “go east” to find reward (Fig. 1A). Contingencies changed either rarely, after 17/20 consecutive correct trials (LFR), or rapidly, after five of seven consecutive correct trials (HFR). Overall performance was better in LFR than HFR (Fig. 1B, left, t test, $t_{(27)} = 3.26$, $p < 0.01$), whereas the number of correct trials was similar (Fig. 1B, right). Although all rats learned the LFR task before the HFR task, tetrode positions remained within $\sim 100 \mu\text{m}$ across the tasks (Fig. 1C) and neurons recorded during LFR and HFR fired at equivalent rates (Fig. 1D, 2-way ANOVA, $F_{\text{task}(1,367)} = 1.8$, $p > 0.05$) and with similar distributions on the maze (Fig. 1E, ks test, $p \approx 1$).

Reversal frequency altered predictive, but neither reward site nor retrospective coding

Reward-site signals

The two goal arms defined distinct spatial stimuli that predicted the same chocolate sprinkles reward. If OFC firing rates vary only with expected reward value or outcome identity, then unit activity should be similar in the two goals. If the OFC signals stimulus–outcome associations, however, then unit activity should distinguish the two goals. We therefore compared the firing rate of each neuron while the rats occupied the east and west reward sites during correct trials before reward consumption (Fig. 2*Ai*, reward-site). Reward site selectivity *p*-values corresponded to effect size (*p* level, proportion with Cohen's *d* value: LFR, $p < 0.05$, $d > 0.8$, 46%, $d > 0.5$, 97%; HFR, $p < 0.05$, $d > 0.8$, 55%, $d > 0.5$, 100%). Although the reward identity and quantity were equivalent, ~25% of the neurons fired with significantly different rates in the two goals, discriminating specific stimulus–outcome associations equally in both tasks (Fig. 2*B–Di*; reward site: LFR, 27% vs HFR, 23%, $\chi^2 = 0.30$, $p = 0.58$).

Retrospective signals

OFC cells fire in patterns that differ statistically during identical spatial trajectories depending on the past or future behavior of the animal (Young and Shapiro, 2011b). In this experiment, the same goal arm was entered from two different start arms, so the same expected outcome was predicted by two different stimuli. If OFC activity includes retrospective information, then firing rates in the goal arm should distinguish the two start arms (Fig. 2*Aii*). Indeed ~25% of the neurons fired at different rates in the goal arms depending on the start of the journey (Fig. 2*B–Dii*, “Retrospective”) and effect size corresponded with *p*-values (*p* level, proportion with Cohen's *d* value: LFR, $p < 0.05$, $d > 0.8$, 91%, $d > 0.5$, 100%; HFR, $p < 0.05$, $d > 0.8$, 94%, $d > 0.5$, 100%). Moreover, each goal arm was associated with equivalent numbers of rewards in the LFR and HFR tasks and those rewarded journeys were equally divided between the two start arms. If OFC firing rates vary with reward number, then the proportion of units with retrospective activity in the goal arms should be equivalent in both tasks. Indeed, the proportion of OFC units with discriminative firing in the goal arms was independent of the rate of contingency changes (Fig. 2*B–Dii*, “Retrospective”: LFR, 25% vs HFR, 23%, $\chi^2 = 0.09$, $p = 0.76$). After the choice point, OFC firing rates distinguish stimulus–outcome associations but not the stability of contingencies.

Predictive signals

OFC firing patterns in the start arm that predict which goal will be chosen imminently could reflect outcome expectancy, a representation of the reward location associated with start arm (Fig. 2*Aiii*). OFC units distinguished the goal of journeys in the start arms by firing at significantly different rates in both tasks and the magnitude of start arm selectivity corresponded with effect size (*p* level, proportion with Cohen's *d* value: LFR, $p < 0.05$, $d > 0.8$, 100%; HFR, $p < 0.05$, $d > 0.5$, 100%, $d > 0.8$, 88%). If start arm activity indeed reflects expected outcomes and forming these expectations depends upon relatively stable contingencies, then predictive coding should be more common in LFR than in HFR tasks. Predictive coding was more common in the LFR task, when reward contingencies were relatively stable, than in the HFR task, when contingencies changed more rapidly even as movement variables were indistinguishable (predictive LFR, 19% vs HFR, 7%, $\chi^2 = 5.88$, $p = 0.013$; Fig. 2*B–Diii*). Before rats reached the choice point, OFC firing rates predicted choices most often when contingencies were stable, showing that OFC signals associated

with the same reward distinguish identical spatial paths taken toward different goals. Only predictive coding was affected by the stability of contingencies, the task epoch in which OFC activity could inform the pending choice, perhaps by signaling outcome expectancy.

Predictive coding varied with reward history, not ongoing performance levels

Although rats earned the same total number of rewards, overall performance differed in the LFR and HFR tasks and in reward stability and either factor could affect OFC activity. If predictive OFC signals reflect reward history integration, then they should vary with the stability of contingencies more than overall or instantaneous performance levels. If predictive signals reflect overall or present performance, however, then unit activity should vary with performance levels independently of the history of contingency changes. We compared these hypotheses using two analyses to quantify OFC coding as rats performed at identical levels in the LFR and HFR tasks.

The first analysis compared a subset of full LFR and HFR sessions matched for performance (Fig. 3*A*, top). Predictive coding was more common in LFR than HFR when performance levels were statistically indistinguishable (Fig. 3*A*, “Coding”: LFR, 24% vs HFR, 7%, $\chi^2 = 6.68$, $**p = 0.009$) and the proportion of predictive cells was similar in the performance matched and full set of sessions (Fig. 2*Diii*). Retrospective coding was also equally likely in the performance matched and full set of sessions, a positive statistical control (~25%; cf. Figs. 2*Dii*, 3*A*). The second analysis included all sessions and compared all HFR trials with performance matched blocks of truncated LFR trials (Fig. 3*B*, top). To match the performance of the HFR trials (5/7 reversal criterion), each block of LFR was truncated to include the first 10 correct trials within a moving window of 14 trials (10/14 rather than five of seven LFR trials were used to maintain sufficient trial numbers for statistical tests). Predictive coding was equally common in the truncated sessions: more units showed prospective coding in LFR (21/98, 22%) than HFR (7/90, 8%), the proportions differed significantly between the tasks ($\chi^2 = 5.86$, $*p = 0.016$; Fig. 3*B*) and matched the full sessions (cf. Figs. 2*Diii*, 3*A*). As before, the ~25% proportion of OFC units showing retrospective coding was similar across tasks and trial selection methods (cf. Figs. 2*Civ*, 3*A*, *B*). Retrospective signals discriminate the start of behavioral episodes and were unaffected by different reversal frequencies. In summary, whereas contingency stability had no effect on retrospective coding, predictive coding was reduced in HFR when performance levels and total number of rewards were matched. Only predictive codes were affected by the extended history of rewards and varied with the rate of contingency changes, consistent with a representation of expected outcomes.

Memory errors occur for several reasons, including trace decay and miscoding during stimulus presentation or retrieval. In miscoding, neural activity maintains strong discriminative signals for the incorrect option: in trace decay, the discriminative signal diminishes (Hampson et al., 2002). If OFC activity informs choices, then neuronal activity should differ in correct and error trials, for example, by showing weaker discriminative signals. We therefore compared outcome selectivity in LFR and HFR by assessing firing rates during correct and error trials for each journey (Fig. 2*Aiv*). During both LFR and HFR, ~25% of the recorded neurons fired at significantly different rates on correct and error trials before the rat reached the goal and outcome selectivity corresponded with effect size (*p* level, proportion with Cohen's *d*

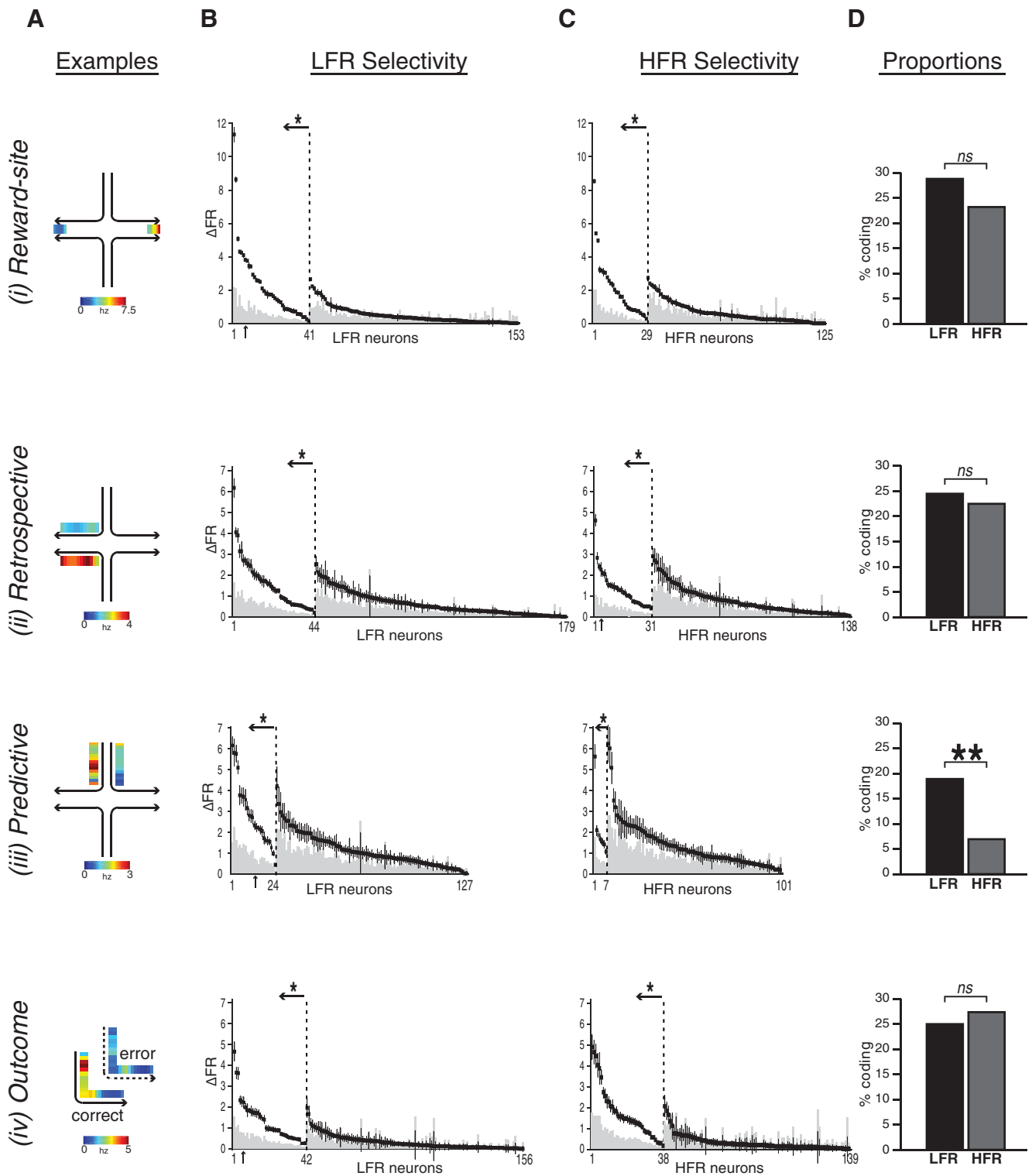


Figure 2. Task-related differences in firing rates. **A**, Single-unit activity examples showing firing rate selectivity (scale shown below each example). **Ai**, Reward-site-selective neuron that fired at higher rates in the east than the west goal. **Aii**, Retrospective coding neuron that fired with higher rates in the west goal arm in journeys from the south start arm compared with journeys from the north. **Aiii**, Predictive neuron that fired at higher rates in the north start arm before the rat entered the west than the east goal arm. **Aiv**, Outcome-elevative neuron that fired at higher rates during correct than incorrect northeast journeys. **B**, **C**, Differential firing rates quantify selectivity. Differences in mean firing rates (vertical axis) are plotted for each eligible unit (horizontal axis) in the LFR (**B**) and HFR (**C**) tasks. Each black square shows the mean firing rate difference of one single unit (vertical lines: \pm SEM) and the gray columns show the mean of permuted firing rate differences. Units with significant selectivity are left of the dashed vertical lines (set off by \leftarrow^* : \uparrow : examples in **A**). **D**, Proportion of reward site, retrospective, and outcome selective neurons were equally common in both tasks; predictive coding was significantly more common in LFR than HFR.

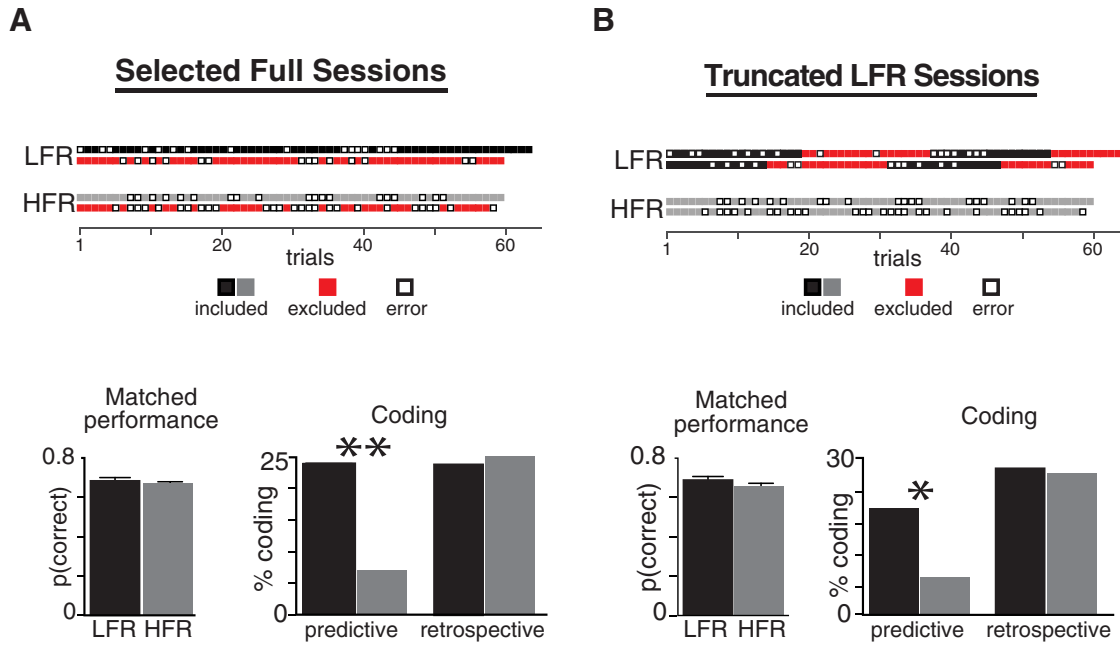


Figure 3. Predictive coding and performance accuracy. Example sessions in the top row use boxes to depict correct (filled) and error (empty) trials in LFR (black) and HFR (gray) sessions; red boxes indicate excluded trials. **A**, Overall mean performance was equivalent in nine complete LFR and 12 HFR sessions (bottom left, mean \pm SEM). In these performance-matched sessions, predictive coding was more common in LFR than HFR, whereas retrospective coding was equally common (bottom right). **B**, Each block in every LFR session was truncated when rats performed 10/14 correct trials (black) to match the performance in full HFR sessions (gray). Predictive coding (left) was more common in LFR than in HFR in these performance-matched sessions, whereas retrospective coding was equally common (right).

value, LFR: $p < 0.05$, $d > 0.8$, 78%, $d > 0.5$, 100% HFR: $p < 0.05$, $d > 0.8$, 61%, $d > 0.5$, 100%; Fig. 2*Biv–D*). Differential activity in correct and error trials could signal uncertainty or confidence about pending reward and correlate, for example, with the total number of rewards independently of contingency stability, which was similar in LFR and HFR (Fig. 1*B*). Alternatively, outcome selectivity could reflect the strength of expectancies, for example, if trace decay causes errors. If outcome coding reflects expectancies that require stable contingencies, then error coding should be stronger in the LFR than HFR tasks, but it was not (LFR, 27% vs HFR, 27%, $\chi^2 = 0.2062$, $p > 0.05$; Fig. 2*Div*). If predictive coding reflects expected outcomes, however, then units that both predict choices and signal errors should be more common with stable contingencies—and they were. The proportion of outcome-selective neurons that also predicted goals were more common in LFR (9/42, 21%) than HFR (1/38, ~2%; Fig. 2), as was the proportion of predictive units that also differentiated outcome (LFR, 9/24, 38%; HFR, 1/7, 16%). In both cases, however, the distribution of conjunctive coding cells matched the proportion of predictive units (predictive vs conjunctive: $\chi^2 = 0.002$, $p = 0.96$), suggesting that the same mechanisms may drive conjunctive goal and outcome prediction coding, for example, by integrating reward history.

OFC ensemble activity predicted learning of stable contingencies

Contingency changes alter PFC unit activity so that population codes become anticorrelated as rats learned to select different goals (Rich and Shapiro, 2009; Young and Shapiro, 2011a). To assess the effects of contingency stability on predictive coding, we analyzed OFC ensemble activity in the start arms as rats approached the choice point during correct trials. We sorted LFR trials into two contingency blocks divided into sub-blocks during initial learning (sub-blocks 1–3; Fig. 4*A*, left) and early, middle,

late, and stable performance of the new contingency (Fig. 4*A*, left) and sorted HFR trials into four contingency blocks divided into two sub-blocks using split-half procedures (sub-blocks 1–8; Fig. 4*A*, right). PVs included the mean activity of every unit in each ensemble, were defined for each sub-block, and the correlation between PVRs measured coding dynamics across the learning curve. PVRs measured ensemble transitions by comparing “seed” PVs to the PV of each other sub-block (LFR, r_{first} , r_{final} ; HFR, r_{1-4} ; Fig. 4*A*). Each ensemble included >4 active units and the analyses compared statistically indistinguishable numbers of neurons, ensembles, and neurons per ensemble in the 2 tasks (number of ensembles: LFR, 10, HFR, 11; mean number of units per ensemble: LFR, 10, HFR, 8, $t_{(19)} = 1.4$, $p = 0.17$).

In both tasks, OFC PVs were stable and highly correlated across successive trials before the contingency reversed (mean PVRs before: LFR sub-blocks 1–3, $0.4 < r < 0.8$; HFR sub-blocks 1–2, $0.5 < r < 0.80$) and then changed rapidly after (LFR, $0.2 < r_{\text{first}}$, early < 0.01 ; HFR, $0.1 < r_{\text{first}}$, block 3 < 0.2 ; Fig. 4*A*). Population coding differences between tasks emerged as reversal performance improved. As rats learned a new and stable contingency in LFR, OFC ensembles developed new and distinct firing patterns that predicted performance changes and PVs became increasingly anticorrelated with the stable PVs recorded before the reversal ($r \neq 0$: LFR, r_{first} , block M < -0.2 , $z = 2.18$, $p < 0.05$; r_{first} , block L < -0.35 , $z = 2.6$, $p < 0.01$; r_{first} , block S < -0.45 , $z = 2.74$, $p < 0.01$; Fig. 4*A*). The anticorrelation reflected both the decline of the original state (blue line) and the emergence of a new one (red line) and both types of coding changes correlated with learning (LFR, r_{first} vs performance: $r^2 = 0.83$, $p < 0.05$; r_{final} vs performance: $r^2 = 0.96$, $p < 0.05$). OFC ensembles developed distinct activity patterns that distinguished pending choices when contingencies were stable.

In contrast to LFR, PVs did not become anticorrelated when rats switched between HFR contingencies ($r \neq 0$: mean $r =$

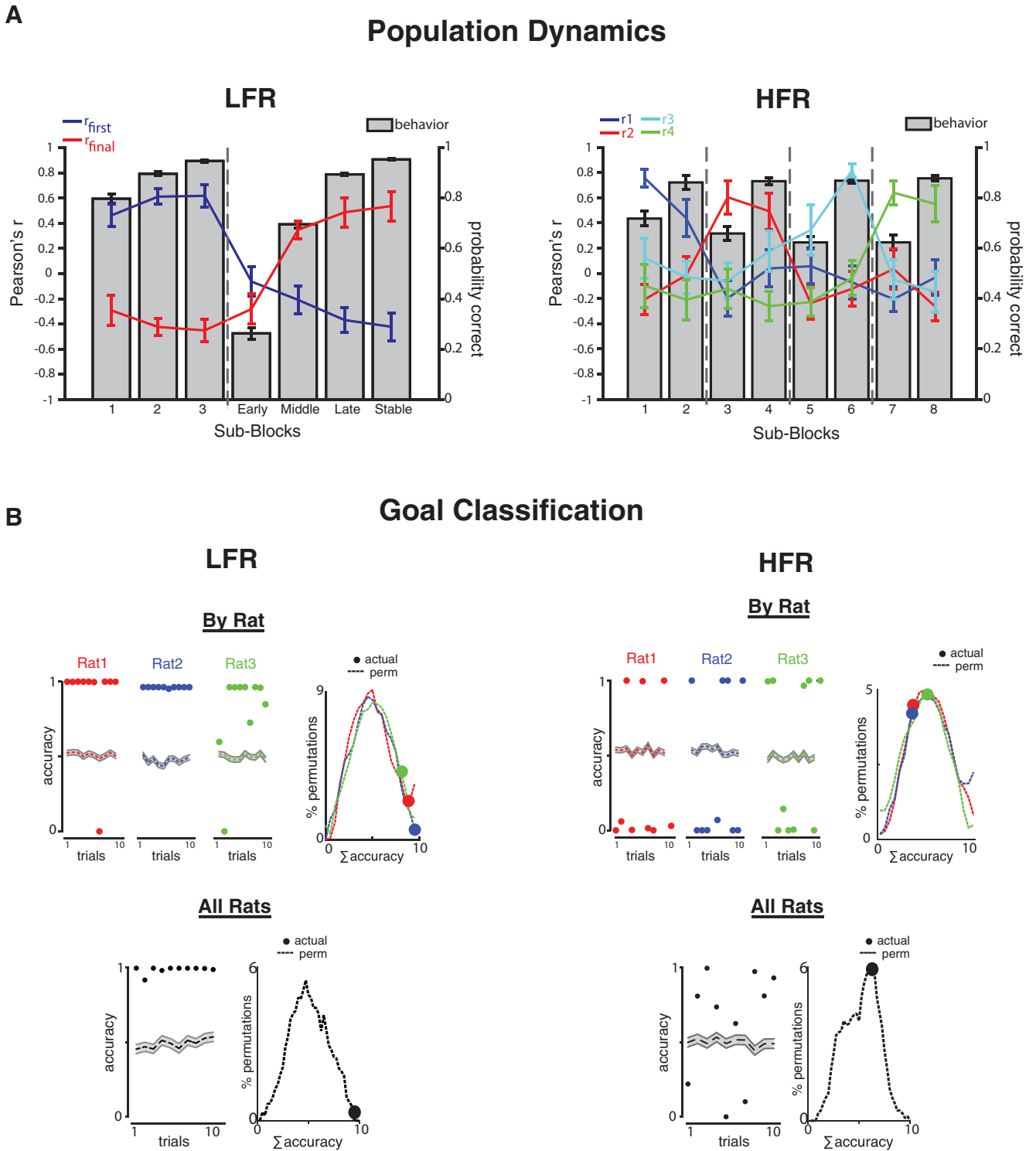


Figure 4. OFC population dynamics and predictive decoding. Firing restricted to the start arms in behaviorally indistinguishable correct trials was used to analyze predictive coding and decoding. **A**, Transitions between activity states were quantified by mean PV correlations of each trial block with a seed PV based on the performance of each contingency block (LFR, r_{first} blue, r_{final} red; HFR, r_1 , r_2 , r_3 , r_4). The colored lines show the mean Pearson's r (\pm SEM) for each trial sub-block (left ordinate); bars plot probability correct (\pm SEM) during the same trials (right ordinate). Abscissa labels indicate trial blocks and dashed vertical lines show contingency changes. LFR trials (left) were grouped into sub-blocks by performance level: blocks 1–3 show stable performance of the initial contingency, early, middle, and late, and stably correspond to performance stages of the new contingency. HFR trials (right) were grouped according split-half procedures within each trial block (sub-blocks 1–8). **B**, Goal classification by population vectors in LFR (left) and HFR (right). Unit activity in the start arm during the first five correct in seven trials during blocks 1 and 2 was compiled into PPVs for each task and rat. Lasso regression models decoded the goal of each of the 10 correct trials and a logistic function determined classification accuracy (where 1 = perfect decoding; 0 = poor decoding). Left, Accuracy ($1 - error^2$) in each of the 10 trials is plotted as circles for each rat (red, blue, and green = rats 1, 2, and 3, respectively) and the accuracy of permutation tests is plotted by colored dotted lines (mean \pm SEM gray). Right, Accuracy distribution of all permutations [vertical axis, $\leq (1 - error^2)$, dotted lines] and the actual accuracy (colored circles) are plotted for each rat. LFR trials were decoded accurately; HFR trials were not. Bottom, Goal decoding by all active neurons in all rats combined was accurate in LFR trials (\sim 99%), but not in HFR trials (\sim 64%).

–0.09, $Z = 0.85$, $p = 0.19$; 1 of 24 correlations differed significantly from 0, as expected by chance; Fig. 4A) and PVR dynamics did not predict HFR learning (r_{1-4} vs performance: $0.21 < r^2 < 0.46$, $p > 0.05$). Furthermore, new OFC firing patterns developed each time contingencies switched so that PVRs comparing OFC activity as rats followed identical contingencies were uncorrelated (e.g., R1 and R3). Rather than representing stable stimulus–reward associations, OFC ensembles developed new ensemble codes after each contingency change as though predictive codes were associated with contiguous episodes rather than abstract rules.

Stable contingencies supported single trial goal decoding

Linear classifiers decode neural activity states associated with different choices or goals (Rich and Wallis, 2016). To determine whether prospective goal decoding differed in LFR and HFR tasks, we used OFC ensemble activity recorded in the start arm to predict goal choices as rats learned to choose a new goal correctly in five of seven trials. OFC ensembles predicted goals more accurately in LFR than HFR in each rat considered separately and in all rats taken together. During LFR, ensembles in each rat predicted choices accurately in >83% of all trials, significantly better than chance (rats 1–3: 90%, 99.9%, and 83.14%; $p = 0.04$, $p = 0.002$, $p = 0.053$ in permutation tests, respectively; Fig. 4B, top). In contrast, the number of choices predicted accurately in HFR was not better than chance (37.8%, 35.3%, and 45.2%; $p = 0.84$, $p = 0.79$, $p = 0.56$, respectively; Fig. 4B, top). Similar results were obtained when activity in all three rats was pooled (accuracy: LFR, 98.9%, $p = 0.005$; HFR, 62.4%, $p = 0.249$; Fig. 4B, bottom) and after constraining LFR ensemble size to match HFR ensemble size (100 randomly selected ensembles of 47 units showed decoding ~94% accurate). OFC ensembles predicted goal choices accurately when stable contingencies were maintained across trials.

Discussion

OFC function improves learning when new outcomes conflict with reward expectancies established by stable contingencies and OFC firing rates predict goal choices in similar testing conditions (Young and Shapiro, 2011a). Here, we altered contingency stability by varying the number of rewarded trials between goal reversals in a plus maze and compared unit activity as rats used identical behaviors to approach the choice point on the way to two different spatial goals. When contingencies were stable, the firing rates of single OFC units signaled pending choices, OFC population dynamics correlated with learning, and ensemble activity correctly predicted choices of most single trials. When reward contingencies changed more rapidly, the firing rates of single units were less likely to predict choices, OFC population dynamics did not correlate with learning, and ensembles predicted choices of single trials no better than chance. Reward stability did not affect goal site, outcome, or retrospective selectivity, demonstrating that stable contingencies are required selectively for choice prediction. Together, the results are consistent with the claim that OFC integrates reward history, show that integration requires repeated episodes, and suggest that integration mechanisms are constrained by network dynamics (e.g., the size of incremental learning steps) that associate stimuli with reward to represent expected outcomes.

Contingency stability, reward history, and expectancies

Contingencies define the link between environmental opportunities, actions, and their potential outcomes. When contingen-

cies are stable, common actions in similar circumstances lead to reliable outcomes, accumulated evidence supports stronger prediction, and integrated reward history supports computations that represent expected outcomes. When contingencies are unstable, identical actions lead to unreliable outcomes so that accumulated positive and negative evidence cancels and undermines prediction. In the present experiment, obtaining reward in every correct choice was certain, whereas the rate of reward for a given choice differed between the LFR and HFR tasks. The same contingency manipulations that altered the effects of OFC lesions on learning (Riceberg and Shapiro, 2012) modulated predictive coding by OFC units: OFC lesions impaired LFR but improved HFR and OFC units predicted choices only in LFR. Learning informed by the OFC also established predictive OFC firing patterns and both require relatively stable contingencies.

OFC dynamics

As stimulus–reward associations repeat over sufficient trials, OFC population vectors developed predictive activity that remained stable while the same contingency applies. After a new and stable contingency was established, the same neural ensembles predicted correct choices to the new goal with distinct and anticorrelated PVs. Through the course of reversal learning, however, OFC activity changed in two phases. Almost immediately after a new contingency is imposed, OFC PVs that predict correct choices to the first goal disappeared, replaced by new and uncorrelated firing patterns. This “collapse” (van Wingerden et al., 2010b) of the initial PV was identical in LFR and HFR. In contrast, the formation of new and anticorrelated population vectors occurred more gradually in the LFR task, as would be expected if OFC circuits accumulate reward associations slowly over repeated trials. In LFRs, anticorrelated PVs formed and OFC lesions impaired reversal learning (Riceberg and Shapiro, 2012). The gradual formation of anticorrelated PVs during reversal learning parallels the slow emergence of NMDA-receptor-dependent odor discrimination codes by OFC neurons (van Wingerden et al., 2012) and predicts faster olfactory reversal learning (Stalnaker et al., 2006). The combined results suggest that stable contingencies may let OFC circuits accumulate reward associations over repeated trials, integrate reward history into expected outcomes, and speed learning by providing discriminative error signals.

PVs were almost never anticorrelated in HFR (Fig. 4), perhaps because contingencies changed too quickly for accumulated stimulus–reward associations to separate OFC representations fully. Moreover, OFC representations differed even when behavioral episodes were supported by the same contingency. PVs associated with one contingency were uncorrelated after two reversals, for example, when “go east” blocks were separated by a “go west” contingency (Fig. 4A, r1 vs r3; r2 vs r4), suggesting that OFC representations may be tied more closely to reward in proximal episodes than to more generalized rules. Future experiments should determine how repeated stable contingencies affect OFC representations.

OFC dynamics may account for the paradoxical improvement in HFR learning after OFC lesions (Riceberg and Shapiro, 2012). When contingencies change frequently, predictive coding is weak and population codes associated with each goal remain uncorrelated. If OFC activity nonetheless influences reversal learning, then poorly separated OFC signals would likely increase interference, for example, by reducing hippocampal prospective coding, and thereby slow reversal learning. From this view, OFC lesions may reduce interference and allow other circuits to inform

choices, for example, via “recency-weighted” rewards (Walton et al., 2010).

Learning by distributed networks

Many brain networks respond to changing contingencies and their interactions affect learning and memory (Kim and Baxter, 2001). Functional interactions between the OFC and the ventral tegmental area (VTA), ventral striatum, dorsal raphe nucleus, and the basolateral nucleus of the amygdala (BLA) could provide mechanisms that link stable reward to OFC and support predictive coding. VTA error prediction signals depend on OFC (Takahashi et al., 2011), optogenetic excitation of the dorsal raphe increases prospective coding in the OFC (Zhou et al., 2015), and striatal serotonergic manipulations reverse compulsivity caused by OFC damage (Schilman et al., 2010). OFC and BLA interactions are especially relevant to reward associations, the effects of OFC lesions on learning, and the present results linking contingencies to predictive coding. OFC lesions reduce flexible, outcome related responses in BLA neurons (Saddoris et al., 2005) and BLA lesions reduce OFC outcome expectancy signals (Schoenbaum et al., 2003). Moreover, reversal learning impaired by OFC lesions is rescued by adding BLA lesions (Stalnaker et al., 2007). This paradoxical effect suggests that the combined lesion reduces interference in other circuits that support reversal learning, perhaps by eliminating spurious reward predictions. OFC lesions may improve rapid reversal learning (Riceberg and Shapiro, 2012) by the same mechanism, facilitating intact circuit function by removing noisy or unreliable outcome expectancy signals.

Different tasks and species

Observed OFC activity patterns vary with task parameters, experimental design features, analytic methods, and species (Boulougouris et al., 2007; Butter et al., 1969; Kennerley and Wallis, 2009; Mar et al., 2011; Meunier et al., 1997; Rich and Wallis, 2016). OFC function is often investigated in terms of stimulus–reward mapping using time locked stimulus and reward presentation, for example, odor sampling, waiting, reward consumption epochs. Although behavior is more extended in space and time in the plus maze, the start arm, goal arm, and reward site of the task correspond operationally to features used in stimulus-based tasks. OFC neurons are engaged during extended actions (van Wingerden et al., 2014; Young and Shapiro, 2011b), but the extent to which a place can be considered a stimulus is less clear. In both spatial and nonspatial tasks, however, predictive OFC population activity is acquired relatively slowly and follows similar learning dynamics, for example, when rats perform olfactory go/no-go reversals (van Wingerden et al., 2012), suggesting that both kinds of tasks engage similar OFC computations. Although primates are rarely tested in spatial navigation tasks and their OFC neurons are seldom observed to encode spatial variables (Zald, 2006), analogous firing dynamics have been reported in monkeys. As in the present experiment, reward prediction by OFC neurons was acquired relatively slowly and the proportion of predictive neurons varied with contingency stability in monkeys learning an oculomotor reversal task (Kobayashi et al., 2010). Predictive coding dynamics by OFC neurons in primates also varied with outcome valence. Nearly equivalent proportions of OFC neurons responded differently to positive and negative outcomes when monkeys associated visual stimuli with either liquid reward or an air puff punishment (Morrison et al., 2011). After reversals, reward- and punishment-selective OFC neurons responded maximally to contingency reversals within ~17 trials

(cf. Fig. 5A,B in Morrison et al., 2011), similar to the number of LFR trials needed for predictive coding here.

Theory reconciliation

Fast and slow plasticity mechanisms may help to reconcile two theoretical mechanisms proposed to account for the effects of OFC lesions on behavioral flexibility. One view suggests that the OFC associates stimuli and outcomes rapidly and represents stimulus–reward mappings that provide flexible, “online” reward predictions (Rolls, 2004). A more recent view suggests that the OFC integrates the history of stimulus–reward associations to compute outcome expectancies—representations of the hedonic value and quality of rewards evoked by conditioned stimuli (Schoenbaum et al., 2009; Stalnaker et al., 2015). The present results show that OFC firing patterns respond to new contingencies with both slow and fast dynamics, so that predictive codes emerge slowly as the same contingency holds for a sufficient number of repeated trials and are lost rapidly when contingencies change. In theoretical terms, stable stimulus–reward associations may be integrated slowly to form expected outcomes, but reset rapidly by changing contingencies via stimulus–reward remapping. Although more extensive training in rapid reversals might let OFC populations follow rapid contingency changes with corresponding switches between anticorrelated activity states, lesion experiments suggest otherwise. OFC activity tends to support reversal learning transiently, speeding learning when a well established contingency is altered for the first time (Schoenbaum, 2002; Riceberg and Shapiro, 2012). Other brain regions, such as the medial prefrontal cortex, may switch more rapidly between pattern-separated stimulus–reward representations (Spellman et al., 2015).

References

- Baxter MG, Parker A, Lindner CCC, Izquierdo AD, Murray EA (2000) Control of response selection by reinforcer value requires interaction of amygdala and orbital prefrontal cortex. *J Neurosci* 20:4311–4319. [Medline](#)
- Boulougouris V, Dalley JW, Robbins TW (2007) Effects of orbitofrontal, infralimbic and prelimbic cortical lesions on serial spatial reversal learning in the rat. *Behav Brain Res* 179:219–228. [CrossRef Medline](#)
- Butter CM, McDonald JA, Snyder DR (1969) Orality, preference behavior, and reinforcement value of nonfood object in monkeys with orbital frontal lesions. *Science* 164:1306–1307. [CrossRef Medline](#)
- Hampson RE, Simeral JD, Deadwyler SA (2002) Keeping on track: firing of hippocampal neurons during delayed-nonmatch-to-sample performance. *J Neurosci* 22:RC198. [Medline](#)
- Hastie T, Tibshirani R, Friedman J (2009) *The elements of statistical learning*, Vol 18. New York: Springer.
- Kennerley SW, Wallis JD (2009) Encoding of reward and space during a working memory task in the orbitofrontal cortex and anterior cingulate sulcus. *J Neurophysiol* 102:3352–3364. [CrossRef Medline](#)
- Kim JJ, Baxter MG (2001) Multiple brain memory systems: the whole does not equal the sum of its parts. *Trends Neurosci* 24:324–330. [Medline](#)
- Kobayashi S, Pinto de Carvalho O, Schultz W (2010) Adaptation of reward sensitivity in orbitofrontal neurons. *J Neurosci* 30:534–544. [CrossRef Medline](#)
- Mar AC, Walker AL, Theobald DE, Eagle DM, Robbins TW (2011) Dissociable effects of lesions to orbitofrontal cortex subregions on impulsive choice in the rat. *J Neurosci* 31:6398–6404. [CrossRef Medline](#)
- Meunier M, Bachevalier J, Mishkin M (1997) Effects of orbital frontal and anterior cingulate lesions on object and spatial memory in rhesus monkeys. *Neuropsychologia* 35:999–1015. [CrossRef Medline](#)
- Morrison SE, Saez A, Lau B, Salzman CD (2011) Different time courses for learning-related changes in amygdala and orbitofrontal cortex. *Neuron* 71:1127–1140. [CrossRef Medline](#)
- Pickens CL, Saddoris MP, Setlow B, Gallagher M, Holland PC, Schoenbaum G (2003) Different roles for orbitofrontal cortex and basolateral

- amygdala in a reinforcer devaluation task. *J Neurosci* 23:11078–11084. [CrossRef Medline](#)
- Riceberg JS, Shapiro ML (2012) Reward stability determines the contribution of orbitofrontal cortex to adaptive behavior. *J Neurosci* 32:16402–16409. [CrossRef Medline](#)
- Rich EL, Shapiro M (2009) Rat prefrontal cortical neurons selectively code strategy switches. *J Neurosci* 29:7208–7219. [CrossRef Medline](#)
- Rich EL, Wallis JD (2016) Decoding subjective decisions from orbitofrontal cortex. *Nat Neurosci* 19:973–980. [CrossRef Medline](#)
- Rolls ET (2004) The functions of the orbitofrontal cortex. *Brain Cogn* 55:11–29. [CrossRef Medline](#)
- Saddoris MP, Gallagher M, Schoenbaum G (2005) Rapid associative encoding in basolateral amygdala depends on connections with orbitofrontal cortex. *Neuron* 46:321–331. [CrossRef Medline](#)
- Schilman EA, Klavir O, Winter C, Sohr R, Joel D (2010) The role of the striatum in compulsive behavior in intact and orbitofrontal-cortex-lesioned rats: possible involvement of the serotonergic system. *Neuropsychopharmacology* 35:1026–1039. [CrossRef Medline](#)
- Schoenbaum G, Nugent SL, Saddoris MP, Setlow B (2002) Orbitofrontal lesions in rats impair reversal but not acquisition of go, no-go odor discriminations. *Neuroreport* 13(6):885–890.
- Schoenbaum G, Setlow B, Saddoris MP, Gallagher M (2003) Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. *Neuron* 39:855–867. [CrossRef Medline](#)
- Schoenbaum G, Roesch MR, Stalnaker TA, Takahashi YK (2009) A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. *Nat Rev Neurosci* 10:885–892. [CrossRef Medline](#)
- Smith AC, Frank LM, Wirth S, Yanike M, Hu D, Kubota Y, Graybiel AM, Suzuki WA, Brown EN (2004) Dynamic analysis of learning in behavioral experiments. *J Neurosci* 24:447–461. [CrossRef Medline](#)
- Spellman T, Rigotti M, Ahmari SE, Fusi S, Gogos JA, Gordon JA (2015) Hippocampal-prefrontal input supports spatial encoding in working memory. *Nature* 522:309–314. [CrossRef Medline](#)
- Stalnaker TA, Roesch MR, Franz TM, Burke KA, Schoenbaum G (2006) Abnormal associative encoding in orbitofrontal neurons in cocaine-experienced rats during decision-making. *Eur J Neurosci* 24:2643–2653. [CrossRef Medline](#)
- Stalnaker TA, Franz TM, Singh T, Schoenbaum G (2007) Basolateral amygdala lesions abolish orbitofrontal-dependent reversal impairments. *Neuron* 54:51–58. [CrossRef Medline](#)
- Stalnaker TA, Cooch NK, Schoenbaum G (2015) What the orbitofrontal cortex does not do. *Nat Neurosci* 18:620–627. [CrossRef Medline](#)
- Takahashi YK, Roesch MR, Wilson RC, Toreson K, O'Donnell P, Niv Y, Schoenbaum G (2011) Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex. *Nat Neurosci* 14:1590–1597. [CrossRef Medline](#)
- van Wingerden M, Vinck M, Lankelma J, Pennartz CM (2010a) Theta-band phase locking of orbitofrontal neurons during reward expectancy. *J Neurosci* 30:7078–7087. [CrossRef Medline](#)
- van Wingerden M, Vinck M, Lankelma JV, Pennartz CM (2010b) Learning-associated gamma-band phase-locking of action-outcome selective neurons in orbitofrontal cortex. *J Neurosci* 30:10025–10038. [CrossRef Medline](#)
- van Wingerden M, Vinck M, Tijms V, Ferreira IR, Jonker AJ, Pennartz CM (2012) NMDA receptors control cue-outcome selectivity and plasticity of orbitofrontal firing patterns during associative stimulus-reward learning. *Neuron* 76:813–825. [CrossRef Medline](#)
- van Wingerden M, van der Meij R, Kalenscher T, Maris E, Pennartz CM (2014) Phase-amplitude coupling in rat orbitofrontal cortex discriminates between correct and incorrect decisions during associative learning. *J Neurosci* 34:493–505. [CrossRef Medline](#)
- Walton ME, Behrens TE, Buckley MJ, Rudebeck PH, Rushworth MF (2010) Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. *Neuron* 65:927–939. [CrossRef Medline](#)
- Wilson FA, O'Scalaidhe SP, Goldman-Rakic PS (1994) Functional synergism between putative gamma-aminobutyrate-containing neurons and pyramidal neurons in prefrontal cortex. *Proc Natl Acad Sci U S A* 91:4009–4013. [CrossRef Medline](#)
- Wilson RC, Takahashi YK, Schoenbaum G, Niv Y (2014) Orbitofrontal cortex as a cognitive map of task space. *Neuron* 81:267–279. [CrossRef Medline](#)
- Young JJ, Shapiro ML (2011a) Dynamic coding of goal-directed paths by orbital prefrontal cortex. *J Neurosci* 31:5989–6000. [CrossRef Medline](#)
- Young JJ, Shapiro ML (2011b) The orbitofrontal cortex and response selection. *Ann N Y Acad Sci* 1239:25–32. [CrossRef Medline](#)
- Zald DH (2006) The rodent orbitofrontal cortex gets time and direction. *Neuron* 51:395–397. [CrossRef Medline](#)
- Zhou J, Jia C, Feng Q, Bao J, Luo M (2015) Prospective coding of dorsal raphe reward signals by the orbitofrontal cortex. *J Neurosci* 35:2717–2730. [CrossRef Medline](#)