Title

A triple dissociation across the medial, ventral, and lateral orbitofrontal cortex in rats making sequential foraging decisions

Authors

Paul J. Cunningham[1], *A. David Redish[2]

Affiliations

[1]Department of Psychiatry, University of Minnesota, Minneapolis MN, USA.

[2]Department of Neuroscience, University of Minnesota, Minneapolis MN, USA.

Correspondence

*redish@umn.edu

Summary

The orbitofrontal cortex (OFC) is functionally heterogeneous across its medial-to-lateral axis, though the nature of this heterogeneity is unclear. The present study leveraged high-density neural recording across the medial-to-lateral span of the OFC using a neuroeconomic task in rats to clarify how functional heterogeneity within the OFC participates in sequential cost-benefit foraging. Lateral and ventral OFC contained opposing representations of an encountered reward's subjective benefit and its associated opportunity costs, respectively. These representations directly competed to make accept-reject decisions during reward evaluation. If a poor-quality reward was mistakenly accepted, ventral and medial OFC participated in reevaluation processes during change-of-mind decisions. Collectively, these results suggest a triple dissociation in representational content and timing across the medial-to-lateral OFC during sequential foraging.

Keywords

Orbitofrontal cortex, foraging, medial orbital, ventral orbital, lateral orbital, decision-making, neuroeconomics.

Introduction

Animals must regularly decide whether to spend limited resources on a currently available reward or search for better rewards in the future[1–3]. Consider a bear as trout swim past, or a rideshare driver deciding whether to accept a current request. Such sequential foraging paradigms are key to understanding how costs and benefits guide behavior [4–8]. The orbitofrontal cortex (OFC) participates in a wide range of processes that enable sequential cost-benefit foraging decisions[9–13]. Importantly, OFC is anatomically and functionally heterogenous across its medial-to-lateral axis[14–18]. Medial and ventral aspects of the OFC are more strongly connected to cortical and sub-cortical regions implicated in emotional regulation, motivation, and goal-directed actions. In contrast, lateral aspects of the OFC are more strongly connected to sensory and motor regions implicated in stimulus-guided actions[19–22]. However, little is known about how sequential cost-benefit foraging computations map onto medial-to-lateral OFC heterogeneity.

A recent synthesis[15] suggested that lateral OFC contributes to behaviors guided by stimuli that signal rewards[23–30], while medial OFC participates when reward value is dynamically changing and actions require complex valuations based on unobservable information[31–36]. Consistent with this possibility, the few studies directly comparing medial and lateral OFC in rats have found that medial aspects track changes in reward value and are required to make inferences about rewards that aren't observable, whereas more lateral aspects reflect stimulus-guided actions[37,38]. Previous work addressing medial-to-lateral heterogeneity within the OFC typically focused on distinctions between far medial and far lateral aspects[14,16,37–42], but evidence suggests that ventral aspects of the OFC also play distinct computational roles[15]. Specifically, VO participates in behaviors that are temporally distant from reward and require inhibition of prepotent behaviors[43–49]. However, it is unclear how distinct computations across the OFC's full medial-to-lateral span contribute to sequential cost-benefit foraging.

Restaurant Row (Figure 1a) is a neuroeconomic sequential foraging task, in which subjects are given a limited amount of time to forage for consumable rewards[50–57]. Upon encountering a "restaurant", subjects enter an "offer zone" (OZ) where they are informed of the delay to reward for the current offer. Accepting an offer requires entering the "wait zone" (WZ); rejecting it entails moving to the next restaurant. In the WZ, the delay counts down until reward is delivered, but subjects can quit the offer at any time by leaving the restaurant. If subjects wait out the delay in the WZ, reward is delivered and subjects enter the subsequent "linger zone" (LZ), which ends when subjects leave for the next restaurant. When the delay range is small (1-10s), subjects can achieve their long-term consumption goals by accepting every reward. When the delay range increases (1-30s), subjects must adapt to the reward-scarce environment. Subjects initially adapt by learning to quit bad offers in the WZ, and eventually learn to reject bad offers in the OZ rather than entering the restaurant and wasting time waiting before quitting[57]. Thus, sequential cost-benefit decisions on Restaurant Row are defined by an initial valuation in the OZ followed by a subsequent re-evaluation in the WZ. Initial evaluations and subsequent reevaluations are governed by distinct processes within the frontal cortex, striatum, and hippocampus[50,52,55,56]. Restaurant Row therefore provides an ideal platform to study the

diverse computational roles that the medial-to-lateral OFC plays in sequential, cost-benefit foraging decisions.

The present study leveraged high-density neural ensemble recordings of lateral orbital (LO), ventral orbital (VO), and medial orbital (MO) cortices while rats navigated Restaurant Row. We discovered distinct neural dynamics within each OFC subregion. LO reflected decisions to approach and wait for reward by encoding current reward's subjective benefits. In contrast, VO reflected decisions to leave the current restaurant and search elsewhere for reward by representing the opportunity costs incurred by accepting the current offer, thereby supporting decisions to reject rewards that were not worth waiting for. Representations in LO and VO reflected distinct processes during initial valuation and subsequent revaluation of sequential foraging decisions. MO ensembles primarily reflected correcting mistakenly accepted offers during re-evaluation (change-of-mind) decisions. These decision-relevant representations emerged with experience adapting to the reward-scarce environment in Restaurant Row.

Results

Rats made sequential cost-benefit foraging decisions in Restaurant Row

Consistent with previous work, the probability of skipping increased with delay (Figure 1B; ANOVA: $F(1,4279)=2855.01$, $p<.0001$). Accept decisions were faster than skip decisions (Figure 1C; Mann-Whitney U Test; $z=-55.31$, $p<.001$), and delay influenced the speed of accept and skip decisions (ANOVA: main effect of decision, $F(1,6366)=332.81$, $p<.0001$, significant delay×decision interaction, $F(1,6366)=12.15$, $p<.0001$). Short-delay offers were accepted more quickly than long-delay offers (Figure 1d; ANOVA: $F(1,4143)=41.26$, $p=1.49\times10^{-10}$), whereas long-delay offers were skipped more quickly than short-delay offers (Figure 1E; ANOVA: $F(1,2223)=4.76$, $p=.02$). Collectively these results suggest a fast-acting accept process, overridden by a slower skip process, both of which were sensitive to delay. After accepting an offer and entering the WZ, rats rarely quit. Nevertheless, quits were more likely with longer delays (Figure 1F; ANOVA: $F(1,4137)=2137.01$, $p<.0001$) and tended to happen within the first few seconds of WZ entry (Figure 1G), making them both efficient and economic.

Overall, these results suggest that rats made sequential cost-benefit foraging decisions in the OZ and WZ. To address how neural ensembles in LO, VO, and MO differed during these decisions, three silicon probes were implanted across the medial-to-lateral axis of the OFC (one in LO, one in VO, and one in MO; see S1 for histology) while rats (n=7; 3M, 4F) foraged for their daily food ration on Restaurant Row. 1457 LO cells, 2283 VO cells, and 2487 MO cells were collected across 253 sessions.

A decreasing gradient of reward-related activity across the medial-to-lateral OFC

A characteristic neural signature of LO is reward-related activity[58–64], though the prevalence of reward-related activity in VO and MO is unclear. To address this question, we analyzed neural activity in three epochs around reward: reward-anticipation (3-0 s prior to reward), reward-delivery (0-1 s following reward delivery), and reward-consumption (2-3 s following reward delivery). We found a significant main effect of OFC subregion (Mixed Effects: $F(2,921)=35.78$,

$p=1.1\times10^{-15}$), a significant main effect of reward epoch (Mixed Effects: $F(2,921)=12.36$, $p=5.0\times10^{-6}$) and a region×epoch interaction (Mixed Effects: $F(4,921)=4.34$, $p=.002$). We therefore assessed neural activity across the medial-to-lateral OFC separately between these three reward-related epochs.

Prior to reward delivery, LO exhibited a rise in normalized firing rate, which peaked when sensory cues (sound of feeder) signaling imminent reward delivery were presented (Figure 2A). VO exhibited a similar but weaker pattern during reward anticipation and delivery (Figure 2A, Mixed Effects; main effect of region, $F(1,4469)=749.17$, $p=1.2\times10^{-152}$ and significant epoch×region interaction, $F(1,4469)=87.49$, $p=1.3\times10^{-20}$). Interestingly, MO activity was unchanged during reward anticipation and delivery, suggesting a gradient of reward-related activity during anticipation of and in response to reward that was strongest in lateral and weakest in medial aspects of OFC.

We also found a distinct, subregion-dependent activity pattern during reward consumption. VO and MO ensembles dipped below baseline during reward consumption (Figure 2A). Thus, while reward anticipation and delivery were characterized by heightened activity in LO and VO, reward consumption was characterized by depressed activity in VO and MO.

Previous research has found that reward-related activity in VO is modulated by the delay preceding its delivery[65]. Interestingly, reward delay did not influence neural activity in any subregion within the OFC in any of the reward epochs (Figure 2B-D; Mixed Effects: no main effect of delay, $F(1,24723)=2.06$, $p=0.15$, no delay×epoch interaction, $F(2,24723)=0.16$, $p=.85$, and no delay×region interaction, $F(2,24723)=1.83$, $p=0.17$). Thus, information about offer delay was not encoded in OFC activity at the time of reward-receipt.

Reward-related activity within OFC represents specific reward features, such as flavor[66–72]. To explore the degree to which OFC subregions represented specific reward features, we used a Bayesian algorithm[73] to decode representations of reward flavor by restaurant and then reoriented the data relative to the restaurant the rat was currently at (Figure 2E). Ensembles within each OFC subregion encoded reward-specific information during anticipation, delivery, and consumption (Figure 2F-H; S2A-C). Encoding of reward-specific information was significant across subregions prior to reward delivery (Wilcoxon Signed-Rank Test: LO; $z=10.66$, $p=1.6\times10^{-26}$. VO; $z=9.32$, $p=1.1\times10^{-20}$; MO; $z=12.93$, $p=3.2\times10^{-38}$; Mann-Whitney U Test on real vs. shuffled distributions: LO; $z=5.26$, $p=1.4\times10^{-7}$ VO; $z=4.9$, $p=1.1\times10^{-5}$ MO; $z=7.82$, $p=5.2\times10^{-15}$), immediately after reward delivery (Wilcoxon Signed-Rank Test: LO; $z=11.12$, $p=9.4\times10^{-29}$. VO; $z=8.77$, $p=1.7\times10^{-18}$; MO; $z=12.88$, $p=5.6\times10^{-38}$; Mann-Whitney U Test on real vs. shuffled distributions: LO; $z=8.60$, $p=8.1\times10^{-18}$ VO; $z=5.27$, $p=1.4\times10^{-7}$ MO; $z=10.10$, $p=6.6\times10^{-24}$), and during reward consumption (Figure 2c-e; S2a-c; Wilcoxon Signed-Rank Test: LO; $z=13.42$, $p=4.8\times10^{-41}$. VO; $z=14.31$, $p=1.9\times10^{-46}$; MO; $z=13.67$, $p=1.4\times10^{-42}$; Mann-Whitney U Test on real vs. shuffled distributions: LO; $z=11.30$, $p=1.5\times10^{-29}$ VO; $z=14.22$, $p=6.7\times10^{-46}$ MO; $z=12.78$, $p=2.0\times10^{-37}$). Thus, ensembles across the medial-to-lateral OFC contained reward-specific representations throughout reward epochs despite distinct firing rate dynamics across subregions.

LO represented current offer value whereas VO and MO represented opportunity costs during initial evaluations in the offer zone.

OFC plays a critical role integrating multidimensional information to make cost-benefit decisions[38,74–80]. OZ decisions in Restaurant Row depend on sensory information about reward delay (tone pitch), learned relations between tone pitch and delay, current spatial location, and knowledge about reward-restaurant associations to infer current opportunities relative to potential future opportunities. We found significant differences in neural activity during the OZ across subregions (Fixed Effects: $F(2,12249)=201.98$, $p=4.9\times10^{-87}$), along with a significant interaction between subregions and time-spent in the OZ (Fixed Effects: $F(2,12249)=8.16$, $p=.0002$).

LO ensembles exhibited a burst of activity upon encountering an offer, which began prior to onset of the delay-informative tone and did not contain information about reward delay (Figures 3A, 3D; Fixed Effects: no main effect of delay, $F(1,3535)=.012$, $p=.91$). Thus, LO activity bursts at offer onset did not encode information relevant to making the decision, but instead marked a transition in the state of the world defined by encountering a potentially rewarding opportunity. Interestingly, LO exhibited another related but distinct neural signature in the OZ defined by elevated activity for short delays (Figure 3A). LO tuning curves to reward delay, measured during the putative decision period, revealed a non-linear relation between neural activity and reward delay reminiscent of functions that relate subjective benefit to reward delay[81,82] (Figures 3E; Mixed Effects model with non-linear link: $F(1,3417)=23.3$, $p=1.4\times10^{-6}$). Consistent with the notion that LO encoding is derived in part from environmental cues, LO OZ activity was aligned to tone presentation (Figure 3G).

In contrast, VO activity steadily increased with time-spent in the OZ, ramping more steeply for long than short delay offers (Figure 3B). To quantify how VO ramping depended on the interaction between offer value and time-spent in the OZ, we measured the slope of neural activity as a function of time-spent for each delay. Activity ramp steepness increased with offer delay (Figure 3F; Fixed Effects: $F(1,2848)=45.28$, $p=2.0\times10^{-11}$), accelerating faster for longer delay offers. Cell-averaged VO tuning curves to delay during the putative decision period revealed a positive relation between activity and delay (Figure 3E; Mixed Effects, linear link, $F(1,2845)=137.95$, $p=3.9\times10^{-31}$). Unlike LO, VO activity was unaffected by tone presentations (Figure 3G).

A critical decision variable in sequential foraging is the opportunity costs associated with a potential reward[1,83–86], which reflect the unexploited reward lost by pursuing a current reward. In Restaurant Row, opportunity costs can be estimated based on the average expected delay to reward. Thus, VO ramping activity represented the opportunity costs lost by waiting for the current reward instead of leaving to find future rewards.

MO activity qualitatively resembled VO activity throughout the OZ but differed in degree and timing. MO activity ramped as time in the OZ unfolded, more steeply for longer delays (Figures 3C, 3F; Mixed Effects: $F(1,2908)=10.62$, $p=.001$). However, the degree to which MO ramping activity accelerated with delay was weaker than VO (Mixed Effects: no main effect of

region, F(1,5756)=0.51, p=.48, significant delays×regions interaction, F(1,5756)=4.33, p=.04). In addition, MO ramping was better aligned to the moment of leaving rather than entering the OZ (Figure 3C). MO tuning curves during the putative decision period revealed a linear increase in neural activity with delay (Figure 3E; Mixed Effects: F(1,2906)=58.75, $p=2.3\times10^{-14}$), though this effect was weaker than VO (Mixed Effects: main effect of region, F(1,5752) =58.70, $p=2.1\times10^{-14}$, and an interaction, F(1,5752)=37.57, $p=9.4\times10^{-10}$). Finally, similar to VO, MO activity was unaffected by tone presentations in the OZ (Figure 3G). Thus, MO represented opportunity costs associated with waiting for the current reward during the OZ, though this representation was weaker than it was in VO and appeared later in the OZ.

Delay-dependent activity suggested distinct neural dynamics between accept and skip decisions within the OFC. Indeed, distinct activity profiles were apparent both soon after OZ entry (S3A-C; Mixed Effects: main effect of region, F(1,22979)=115.65, $p=1.1\times10^{-15}$, decisions×regions interaction, F(1,22979)=5.54, p=.004, and a 3-way time×decisions×regions interaction F(1,22979)=18.56, $p=6.6\times10^{-9}$) and lasting until OZ exit (S3A-C; Mixed Effects: main effect of region, F(1,22788)=6.93, p=.0009, decisions×regions interaction, F(1,22788)=34.64, $p=9.4\times10^{-16}$, and time×decisions×regions interaction F(1, 22788)=5.52, p=.004). Of particular interest was the opposing representations of subjective benefits in LO apparent early in the OZ and opportunity costs in VO that emerged later. These opposing representations suggest that LO reflected a process that generated quick accepts whereas VO reflected a separate, slower process that overrode the prepotent tendency to approach reward and instead search elsewhere.

Consistent with this hypothesis, there were decision-dependent differences in LO activity relative to VO activity as OZ time unfolded. Early in the OZ, LO activity exceeded VO activity (Figure 4A, 4B, 4E), regardless of which decision was made (Mann-Whitney U Test: comparing LO-to-VO activity at zone entry: accepts: z=3.8, $p=1.3\times10^{-4}$; skips: z=2.19, p=.02). If the offer was accepted, increased activity in LO relative to VO was maintained until rats began running towards the WZ. If the offer was skipped, VO activity overtook LO activity towards the end of the OZ (Figure 4C, 4D, 4F; Mann-Whitney U Test: comparing LO-to-VO activity during skips at zone exit: z=-2.40, p=0.02). As a result, accept decisions were associated with a greater proportion of time-spent in the OZ with LO activity exceeding VO activity compared to skips (S3D). In addition, this effect was strongest when time spent making the decision was relatively short (S3D), which presumably reflected relatively fast accept decisions driven by heightened LO activity.

Collectively, these results suggest that LO and VO reflected opposing processes — LO represented the subjective benefits of the current offer as part of a fast process that promoted approach (accepts). VO, and to a lesser extent MO, represented opportunity costs as part of a slower process that promoted leaving for other opportunities (skips). Importantly, the relative timing of decision- and delay-dependent activity in the LO and VO closely mirrors decision- and delay-dependent reaction times (Figure 1C-E).

In the wait zone, LO represented time to reward while VO and MO contained complementary representations of time-remaining and time-spent, respectively, during re-evaluative quits.

The wait zone (WZ) is a dynamically changing epoch in which temporal progression towards reward was signaled by a descending sequence of tones, providing information about how much time remained until reward. Such future-oriented information is distinct from past-oriented information about how much time has already elapsed in the WZ. This raises the question of the extent to which OFC ensembles across the medial-to-lateral axis encoded information about time-spent, time-remaining, or both.

Given the structure of the WZ countdown, each offer of a given initial delay was associated with a unique trajectory through the two-dimensional space of time-remaining and time-spent. To visualize this, we plotted ensemble tuning curves as a function of time-remaining, time-spent, and initial delay (Figures 5A-C).Each OFC subregion revealed distinct activity profiles to WZ temporal parameters (Mixed Effects: main effect of region, $F(2,1.5\times10^{5})=167.77$, $p=1.7\times10^{-177}$; main effect of time-remaining, $F(1,1.5\times10^{5})=136.93$, $p=1.3\times10^{-31}$, main effect of time-spent, $F(1,1.5\times10^{5})=179.55$, $p=6.5\times10^{-41}$, a time-remaining×time-spent interaction, $F(1,1.5\times10^{5})=129.57$, $p=5.3\times10^{-30}$, and a 3-way interaction of region×time-remaining×time-spent, $F(2,1.5\times10^{5})=17.11$, $p=3.7\times10^{-8}$).

LO ensembles represented both time-spent and time-remaining (Figures 5A, 5D, 5E; Mixed Effects: significant time-spent×time-remaining interaction, $F(1,47493)=75.35$, $p=4.1\times10^{-18}$; with significant main effects of time-spent, $t(1,47493)=10.69$, $p=1.2\times10^{-26}$, and time-remaining, $t(1,47493)=-6.5\times10^{11}$). Representations of time-remaining is consistent with previous work showing that LO responds to cues that predict impending reward[58–62]. In Restaurant Row, this phenomenon appeared as progressively heightened activity as reward approached (Figures 5A, 5D). Sensitivity of LO activity to time-remaining was driven primarily by cue presentations, as evidenced by rhythmic LO bursting time-locked to tone presentations in the WZ (Figure 5F).

VO ensembles represented time-remaining but not time-spent (Figures 5B, 5D, 5E; Mixed Effects: significant time-spent×time-remaining interaction, $F(1,38189)=37.45$, $p=9.4\times10^{10}$, with a significant main effect of time-remaining, $t(1,38189)=-7.9$, $p=1.8\times10^{-15}$, but not time-spent, $F(1,38189)=0.15$, $p=0.87$). This result is perhaps not surprising given the observation that LO and VO exhibited enhanced activity during reward anticipation (Figure 2A). However, here we found that both VO and LO encoded information about time-remaining to reward throughout the entire WZ, rather than during the few seconds preceding reward only. Finally, as in LO, VO representations of time-remaining in the WZ were defined by rhythmic activity in which normalized firing rate peaked by the end of tone presentations (Figure 5F).

MO ensembles represented time-spent but not time-remaining (Figures 5C, 5D, 5E; Mixed Effects: significant time-spent×time-remaining interaction, $F(1,38955)=19.49$, $p=1.0\times10^{-5}$, with a significant main effect of time-spent, $F(1,38955)=-4.72$, $p=2.3\times10^{-6}$, but no main effect of time-remaining, $F(1,38955)=-0.73$, $p=0.47$). Specifically, MO activity was elevated immediately upon entering the WZ, thereby bridging the gap between OZ exit and WZ entry. Unlike LO and VO, MO activity during the WZ was not temporally locked to cue presentations (Figure 5F).

Overall, these results suggest a gradient of cue-driven representations of time to reward, with representations apparent in LO and VO but not MO. Further, whereas LO independently represented both time-spent and time-remaining, VO only represented time-remaining and MO only represented time-spent. Complementary representations of time-remaining and time-spent in VO and MO, respectively, provide the information needed to make re-evaluative quit decisions in the WZ. We thus compared neural activity, oriented with respect to WZ entry and exit time, within each subregion separated based on whether the rat quit or waited to earn reward and found marked differences in the temporal dynamics of neural activity across subregions (Mixed Effects: main effect of region, $F(2,12014)=9.62$, $p=6.6\times10^{-5}$; region×time interaction, $F(2,12014)=5.41$, $p=.004$) and prior to quitting (Fixed Effects: main effect of region, $F(2,12394)=492.72$, $p=1.2\times10^{-206}$; but no region×time interaction $F(2,12394)=1.30$, $p=0.27$).

LO ensembles exhibited a burst of activity soon after entering the WZ (Figure 6A). Given that LO was active during OZ approach decisions and represented time-remaining to reward, we hypothesized that the LO WZ entry burst would be disrupted for encounters that were quit. Interestingly, however, we found the opposite result — LO activity at WZ entry was higher during quit encounters (Figure 6A; Mann-Whitney U Test: $z=-3.31$, $p=9.4\times10^{-4}$). This seemingly paradoxical finding suggests that anomalously high LO activity might result in an erroneous accept decision. Consistent with this hypothesis, LO activity for offers with longer delays were significantly higher just prior to leaving the OZ when the offer would be quit compared to when it was either earned or properly skipped (S4D-E: ANOVA: main effect of decision, $F(2,2765)=7.70$, $p=4.6\times10^{-4}$; post-hoc comparisons revealed significantly higher activity during quits than both skips, $p=.002$, and earns, $p=.002$). Thus, heightened LO activity during quits further reflected the critical role that LO likely plays in accept decisions.

Aligned to WZ exit time, VO activity was elevated in the period leading up to a quit decision. Elevated VO activity during quits exceeded its reward-anticipatory activity when earning reward (Figure 6D; Mann-Whitney U Test: prior to WZ exit, $z=-5.6$, $p=2.01\times10^{-8}$). To determine when this elevated activity emerged, we compared VO activity between stays and quits oriented to WZ entry time. Remarkably, VO already discriminated stay from quit decisions upon entering the WZ. VO activity was elevated during late periods of the OZ that persisted through WZ entry. In contrast, if reward was earned, VO activity dipped below baseline levels (Figure 6C; Mann-Whitney U Test: comparing mean activity between quits and earns at WZ enter, $z=-3.66$, $p=2.52\times10^{-4}$). Decision-encoding in VO was observed when controlling for reward delays (S4B). These data are consistent with quit decisions correcting anomalous LO activity that VO ramping in the OZ did not override.

As noted earlier, MO ensembles bridged the transition between exiting the OZ and entering the WZ (Figure 3C). Given that this activity occurred during the critical period when relatively quick quit decisions were made, we hypothesized that this activity played a role in correcting wrongly accepted offers in the OZ by quitting in the WZ. Consistent with this possibility, MO ensembles differentiated stays from quits during the moments prior to leaving the WZ (Figure 6F). Interestingly, however, this differentiation did not emerge until about 1s

following WZ entry (Figure 6E). Decision-dependent differentiation in MO activity was observed even when controlling for reward delays (S4C).

Collectively, these results highlight that LO, VO, and MO participated in unique ways to re-evaluative change-of-mind decisions. Representations of time-remaining linked to learned cues were present only in LO and VO, whereas representations of time-spent waiting for reward were present only in LO and MO. When re-evaluative quit decisions were made, both VO and MO exhibited elevated activity just prior to quitting, suggesting additional override dynamics.

Neural dynamics in LO and VO reflected distinct processes during initial valuations and subsequent re-evaluations

Representational content and neural dynamics within LO and VO during OZ evaluations and WZ re-evaluations exhibited similarities and differences: LO exhibited bursts when entering each zone and was elevated when reward was temporally proximal. VO showed elevated activity during long delays and was elevated in the moments leading up to decisions that resulted in rats abandoning current rewards via skips or quits. These qualitatively similar dynamics within each subregion might reflect the same process at distinct task stages; however, previous data suggests that distinct decision-making processes drive decisions in the OZ and WZ[50,51,54–57].

To test this, we examined the degree of overlap in the cells participating in key neural signatures within LO and VO throughout task states. For LO, we assessed the overlap in cells participating in bursts aligned to OZ entry, WZ entry, and reward delivery. For VO, we assessed the overlap in cells participating in leaving a task zone during skips, quits, and leaving the restaurant after earning reward.

There was little overlap in LO cells and little correlation in firing rates during OZ and WZ entry bursts (S5D; OZ-sensitive cells: Pearson's Correlation: $r=.13$, $p=.063$; WZ-sensitive cells: Pearson's Correlation: $r=-.11$, $p=.92$). Interestingly, there was a negative correlation in activity scores at OZ entry and reward delivery (S5E; $r=-5.9$, $p=1.9\times10^{-14}$), suggesting that OZ cells tended to be depressed at reward delivery. We did not find a significant correlation in activity at OZ entry and reward delivery for cells sensitive to reward delivery (S5E; Pearson's Correlation: $r=-.04$, $p=.61$). Finally, there was no significant correlation in activity at WZ entry and reward delivery for either WZ-sensitive cells (Pearsons Correlation: $r=.05$, $p=.27$) or reward-delivery-sensitive cells (S5F; Pearsons Correlation: $r=.006$, $p=.46$). This negative correlation was also evident in the peri-event averages, which revealed that cells that were sensitive to a given zone entry were only weakly modulated (if at all) at the other two zone entries (S5A-C). Thus, distinct clusters of LO cells marked transitions into the OZ, WZ, and reward delivery. This result is consistent with prior observations of distinct LO cells marking state transitions in appetitive learning tasks[61,62].

While these results suggest LO bursts and zone entry were distinct, it remains possible that LO cells encoding information about reward delay did so similarly between the OZ and WZ. By comparing shuffle-normalized correlations in tuning curves to delay in the OZ and time-remaining in the WZ for each cell, we found no significant relation between delay-sensitive activity between the two zones (Wilcoxon Signed Rank Test: real versus shuffled distributions,

z=1.2, p=.25; S5G). Collectively, these results suggest that LO ensembles participated in distinct computational processes throughout task states in Restaurant Row, rather than reflecting a singular process that was reactivated during sequential decisions in the OZ and WZ.

VO activity was heightened not only in the moments leading up to skips and quits, but also in the moments leading up to leaving the linger zone (S5J). We therefore assessed the degree of overlap in the collection of cells that participated in leaving a restaurant via a skip, quit, or lingering. There was no significant correlation in average activity prior to skipping and quitting for cells sensitive to either skips (S5K; Pearsons Correlation: z=-.004, p=.52) or quits (S5K; Pearsons Correlation: z=.04, p=.30). In addition, peri-event averages of normalized activity revealed that quit-sensitive cells exhibited less pronounced activity prior to skipping than skip-sensitive cells (S5H). Conversely, skip-sensitive cells exhibited less pronounced activity prior to quitting than quit-sensitive cells. Finally, we did not find a significant correlation between tuning curves within VO cells to reward delay in the OZ and WZ (S5n; Wilcoxon Signed Rank Test: real versus shuffled distributions: z=-.73, p=.46). Thus, the decision-related processes that VO participated in during OZ skips were distinct from the processes that implemented WZ quits.

Surprisingly, there was substantial overlap in the cells that fired prior to quitting out of the WZ and leaving after lingering (S5I, 5J, and 5M). Average activity of quit-sensitive cells when leaving a restaurant prior to quitting and lingering were significantly correlated (S5M; Pearsons Correlation: r=.47, $p=2.9\times10^{-14}$). Average activity of linger-sensitive cells when leaving a restaurant during lingering and quitting were significantly correlated (S5M; Pearsons Correlation: r=.61, $p=6.4\times10^{-18}$). When comparing neural dynamics of each cell type prior to quitting, activity in cells that were sensitive to leaving a restaurant following post-reward lingering ramped up and peaked in a manner similar to quit-sensitive cells (S5I). Conversely, activity in quit-sensitive cells ramped in the moments leading up to a linger-leave decision, though the degree and peak up ramping was weaker than it was in linger-leave-only cells (S5J). Thus, compared to the distinct collections of VO cells that participated in skips and quits, there was substantially more overlap in the cells that participated in quits and leaving after reward.

Overall, these results are consistent with distinct decision-making mechanisms deployed in the OZ and WZ, with opposing processes implemented in LO and VO. While LO ensembles participated in approach-related behaviors in both zones (i.e., accepts and stays), they did so as part of distinct decision-making processes. VO ensembles participated in leave-related behaviors in both zones (i.e., skips and quits), but they did so as part of distinct decision-making processes. However, there was overlap in the processes participating in quitting an offer (prior to reward) and leaving a restaurant following a lingering period (after reward). Thus, VO ensembles appeared to participate in one process when leaving a restaurant while being spatially distant from reward (OZ skips) and a separate process when leaving a restaurant while being closer to reward (quits and linger-leaves).

Neural dynamics across the OFC developed with behavioral adaptations to a reward scarce environment

Previous work has found that introduction of long delays induces a host of behavioral adaptations to deal with the reward-scarce environment. Given the well-established role of OFC in learning and updating reward-related representations[87–92], we analyzed OFC neural activity across four sequential stages of recording sessions: first 5 sessions post-surgery [1-5s delays], next 5 sessions [1-10s delays], first 7 sessions of the reward scarce environment [1-30s delays], and remaining sessions of the reward scarce environment [1-30s delays]. We found that rats did not stop to consider accept vs. skip decisions in the OZ until stage 3 (Figure 7A; ANOVA: $F(3,251)=117.03$, $p=1.9\times10^{-47}$). Before the reward-scarce environment, rats ran straight through the OZ, whereas in the reward-scarce environment, rats began to evaluate offers in the OZ. Rats also learned to spend less time lingering after earning reward (Figure 7B; ANOVA: $F(3,251)=17.91$, $p=1.5\times10^{-10}$) once the reward-scarce environment was introduced.

Rats did not begin skipping and quitting offers until after the reward-scarce environment was introduced. Rats initially dealt with the reward-scarce environment by accepting every offer but occasionally quitting in the WZ, as indicated by the high number of quits per session but low number of skips during stage 3 (Figure 7C). This tendency to quit was replaced by a tendency to skip in stage 4 (Figure 7C). Finally, we found that the psychometric curve relating skip probability to offer delay emerged with experience. The curve was relatively flat during stage 3, reflecting an insensitivity of OZ decisions to reward delay, but then exhibited a slope by stage 4 (Figure 7D; S6A-B; ANOVA: comparing stages 3 and 4, main effects of delays, $F(1,5242)=1729.07$, $p<.0001$ and stage $F(1,5242)=23.92$, $p<.0001$, and a significant delays×stage interaction, $F(1,5242)=635.54$, $p<.0001$). In contrast, psychometric curves relating quit probability to offer delay in the wait zone appeared immediately when quits began during stage 3 and did not change thereafter (Figure 7E; ANOVA: comparing stages 3 and 4, no main effect of stage, $F(1,5082)=0.02$, $p=0.89$, and no delays×stage interaction, $F(1,5082)=0.22$, $p=0.64$). Thus, whereas cost-benefit decisions in the OZ took time to develop, cost-benefit decisions in the WZ appeared at the start of the reward-scarce environment, consistent with previous work using Restaurant Row[54,57].

Cue-driven representations of future outcomes in the OFC develop with experience, and previous work illustrating this has found these in lateral aspects[87,90,91]. We also found that bursts of LO activity were modulated by experience in the reward-scarce environment (Mixed Effects: main effect of stage, $F(3,529) = 8.62$, $p = 1.3\times10^{-5}$, main effect of zone, $F(2,549)=9.01$, $p=.0001$, and an interaction, $F(6,549)=8.28$, $p=1.3\times10^{-8}$) . LO bursts upon WZ entry and reward delivery were most prominent prior to the reward-scarce environment (stages 1-2) and prior to behavioral adaptations in the reward-scarce environment (stage 3; Figure 7F). WZ and reward bursts were subsequently reduced after behavior adapted (Figure 7F; S6F-G). The reduction in LO activity bursts was accompanied by the emergence of the activity burst at OZ entry (Figure 7F; S6E). Bursts of LO activity at OZ entry did not appear until after behavior adapted to the reward-scarce environment. Overall, these results raise the interesting possibility that bursts of LO activity at cues marking task state transitions backpropagated to the earliest cue that predicted potential reward.

To address the question of how delay-sensitive ramping throughout the OZ in VO and MO developed with behavioral adaptations to the reward scarce environment, we analyzed the degree to which activity ramping changed across stages. We measured the slope of neural activity across time-spent in the OZ changed across the four stages. The ramp developed with experience in the economic crisis, and that it did so similarly between VO and MO (Figure 7G; S6C-D; Mixed Effects: main effect of stage, $F(3,237)=3.10$, $p=.03$, but no main effect of region, $F(1,237)=1.99$, $p=.16$, and no stage×region interaction, $F(3,273)=0.13$, $p=0.94$). Thus, VO and MO developed ramping activity immediately upon introduction to the reward scarce environment, and this ramping grew stronger with behavioral adaptations.

To determine how experience affected the degree to which ramping activity was modulated by reward delay, we compared the change in slope to delay over experience across stages. Both VO and MO developed characteristic delay-dependent ramping in the OZ only after encountering the reward scarce environment (Figure 7H; Mixed Effects: main effect of stage, $F(2,424)=3.25$, $p=.04$, but no main effect of region, $F(1,424)=1.69$, $p=0.19$, and no stage×region interaction, $F(2,424)=0.96$, $p=0.38$). Importantly, delay-dependent ramping did not emerge until after behavior adapted to the reward scarce environment (stage 4 coefficients were significantly different from stage 2, $t(424)=2.50$, $p=.013$, but stage 3 coefficients were not, $t(424)=1.15$, $p=0.25$). This result suggests that delay-dependent ramping activity in VO and MO reflected information about the reward-scarce environment after it was used to make adaptive skip decisions.

Discussion

We found that the OFC was dynamically engaged while rats made sequential cost-benefit decisions in a neuroeconomic foraging task. Importantly, neural dynamics across the OFC exhibited a host of similarities and differences as sequential foraging decisions were made. Table 1 provides a summary of the key observations across subregions during sequential foraging.

Results from the present study suggest that normative models of foraging are implemented at the process level by multiple, conflicting decision-making mechanisms[93]. Previous work in Restaurant Row[50,51,54–57] along with work using sequential foraging tasks[94], similarly suggests that multiple decision-making mechanisms are deployed when animals forage for sequentially-encountered rewards. Importantly, distinct decision-making mechanisms during sequential foraging in Restaurant Row were realized in distinct OFC subregions. Initial evaluations were governed by opposing processes that supported OZ decisions. The fast-acting accept process was reflected in heightened LO activity that encoded subjective benefits associated with the current reward. The skip process was slower, overrode prepotent approach tendencies, and was reflected in ramping activity that scaled with opportunity costs in VO (and to a lesser extent in MO). Interestingly, formal models of foraging suggest precisely these representations to make optimal decisions; the profitability of the current offer should be compared to the opportunity costs associated with pursuing that offer to make an accept/reject decision[1,83–85,95]. Thus, results from the present study suggest that these opposing representations are realized in LO and VO/MO as part of conflicting processes that make evaluations to either accept a reward or continue searching.

OFC dynamics differed between initial evaluations in the OZ and change-of-mind re-evaluations in the WZ. LO encoded the current offer's subjective benefit while rats waited for reward. In contrast, VO and MO activity was heightened in the moments leading up to quitting. However, heightened WZ activity in VO (and MO) during quits was not closely related to OZ ramping activity representing opportunity costs. One possible explanation for this difference is that re-evaluative decisions leveraged different representations and neural dynamics than did initial evaluations. This possibility is supported by the finding that a distinct collection of cells within the VO became active prior to skipping and quitting. However, another possibility is that ramping activity encoding opportunity costs was easier to measure during initial evaluations because initiation of the process was time-locked to presentation of the first cue indicating offer delay in the OZ. During re-evaluations, ramping activity might have been obscured because ramping onset was likely initiated at variable times across encounters.

While most empirical and theoretical treatments of functional heterogeneity within the OFC focus on far lateral and far medial aspects[14,16,37–42], results from the present study revealed a drastic difference between LO and VO. During both initial evaluations and subsequent re-evaluations, LO and VO exhibited opposing neural dynamics, and the degree to which LO and VO activity dominated was directly related to whether rats approached (LO dominant) or abandoned (VO dominant) reward. This result emphasizes the need to consider the OFC's entire medial-to-lateral span when investigating its functional heterogeneity.

There is an ongoing debate regarding cross-species homologies in the OFC[11,96,97]. Although OFC in rodents and primates share anatomical connectivity patterns with other cortical and subcortical regions[97–99], it also differs both anatomically and functionally across species. OFC includes a granular cell layer in primates but not rodents[100]. Abstract representations of subjective value are commonly found in primates making cost benefit decisions[101]; however, such abstract representations are rarely found in rodent studies assessing cost-benefit decisions[102], including the present study.

Results from the present study are inconsistent with research assessing medial-lateral heterogeneity in humans interpreted to reflect preferential processing of reward procurement and loss in medial and lateral OFC, respectively[14,103]. Assuming similarity in medial-to-lateral heterogeneity between primates and rodents, this possibility would suggest that lateral aspects of the OFC would be associated with abandoning reward whereas medial aspects would be associated with approaching reward in Restaurant Row. However, this is the opposite of what we found.

Research exploring medial-lateral heterogeneity in non-human primates suggests that lateral OFC is responsible for assigning credit to actions and stimuli that predict reward whereas medial OFC is responsible for estimating and comparing reward values to implement a decision[41,42]. Interestingly, we found that lateral aspects of the OFC encoded decision-relevant information during the period corresponding to initial evaluations, suggesting a role in guiding behavior rather than learning. In addition, we did not find strong evidence that medial aspects of the OFC participated in making decisions when reward values needed to be inferred and compared during initial evaluations. However, medial OFC did participate in re-evaluative

change-of-mind decisions. This suggests that, to the extent the MO participates in using estimates of reward value to make decisions, it does not do so across all decision processes involved in sequential foraging decisions.

Whether inconsistencies revealed by results from the present study reflect true species differences remains unclear. Direct comparisons between primates and rodents using the same task can provide more firm conclusions regarding species differences and similarities in medial-to-lateral OFC heterogeneity. The Web-Surf task[104–106], which is a version of Restaurant Row modified for humans and accesses similar sequential decision-making processes, could provide a fruitful avenue to directly explore OFC functional heterogeneity across species.

Acknowledgements

We would like to thank Chris Bolt and Ayaka Sheehan for their invaluable contributions to surgery and histology, respectively. In addition, we would like to thank Geoff Diehl and the rest of the members of the Redish lab for helpful feedback and guidance throughout the project. This work was supported by NIH training grant T32DA037183.

Author Contributions

P.J.C and A.D.R designed the study while P.J.C (along with Chris Bolt and Ayaka Sheehan) collected data and supported technical aspects of the project. P.J.C and A.D.R conceived and implemented neural and behavioral analyses. Finally, P.J.C and A.D.R wrote the manuscript.

Declaration of Interests

The authors declare no competing interests.

Main Figure Legends

Figure 1. Rats made economic sequential foraging decisions in Restaurant Row (A) In Restaurant Row, rats make sequential decisions in the “offer zone” (OZ) and “wait zone” (WZ) based on the delay to reward. (B) Mean (±SEM) probability of skipping as a function of delay. (C) Probability density function of time spent making accept (blue) and skip (orange) decisions. (D-E) Cumulative density functions of time spent making accept (D) and skip (E) decisions across delays. Blue = shorter delays; Red = longer delays. (F) Mean (±SEM) probability of quitting as a function of delay. (G) Probability density function of time to quit.

Figure 2. LO, VO, and MO encoded reward-specific information through distinct neural dynamics during reward (A) Mean (±SEM) normalized firing rate within LO, VO, and MO aligned to reward delivery. (B-D) Mean (±SEM) normalized firing rate as a function of offer delay during anticipation (B), delivery (C), and consumption (D). (E) Illustration of decoding analysis, current restaurant = location of the rat. Next, past, and other restaurants rotate with the current location for analysis. (F-H) Mean (±SEM) probability of decoding to each of the four rewards in LO (F), VO (G), and MO (H).

Figure 3. Distinct neural dynamics and representational content in LO, VO, and MO during initial evaluations (A-C) Mean normalized firing for LO (A), VO (B), and MO (C)

aligned to OZ entry (left) and exit (right). Top panels show mean (±SEM) activity across all encounters while bottom panels show mean activity broken up by delay. (D-E). Mean (±SEM) tuning curves of LO, VO, and MO ensembles to offer delay when entering the OZ (D) and during the putative decision period (E). (F) Mean (±SEM) slopes relating normalized firing rate to OZ time, broken up by offer delay for LO, VO, and MO. (G) Mean (± SEM) normalized firing rate aligned to tone presentation at t = 0 and t = -1s. Tone presentations lasted 0.5 s and are represented as filled bars along the top of the plot.

Figure 4. Opposing neural dynamics in LO and VO reflected participation in distinct systems during initial evaluations (A-B) Mean (± SEM) normalized firing rate in LO and VO during accepted encounters aligned to OZ entry (A) and exit. (B). (C-D) Mean (±SEM) normalized firing rate in LO and VO during skip encounters aligned to OZ entry (C) and exit (D). (E-F) Mean (±SEM) difference in normalized firing rate between LO and VO at each moment in time during accept and skip encounters aligned to OZ entry (E) and exit (F).

Figure 5. LO, VO, and MO differentially encoded time remaining and time spent during the wait zone (A-C) Mean tuning curves to time remaining and time spent for LO (A), VO (B), and MO (C). Each pixel in the 3-dimensional space reflects mean firing rate defined by time remaining (z-axis) and time spent (x-axis) given the initial offer delay (y-axis) at that moment. Moving from top to bottom along with z-axis means time to reward is approaching. Moving from back to front on the x-axis means more time has already been spent waiting for reward. (D-E). Mean (±SEM) tuning curves to time remaining (D) and time spent (E) for LO, VO, and MO ensembles. (F) Mean (±SEM) normalized firing rate aligned to tone presentation at t = 0 and t = -1. Tone presentations lasted 0.5 s and are indicated by filled bars along the top of the plot.

Figure 6. Differential contributions of LO, VO, and MO to reevaluative change-of-mind decisions (A-B) Mean (±SEM) normalized firing rate in LO during stay (light lines) and quit (dark lines) encounters oriented with respect to WZ enter time (A) and exit time (B). (C-D) Mean (±SEM) normalized firing rate in VO during stay and quit encounters oriented with respect to WZ entry (C) and exit (D). (E-F) Mean (±SEM) normalized firing rate in MO during stay and quit encounters aligned to WZ entry (E) and exit (F). Note that WZ exit during stay encounters corresponds to reward delivery, whereas WZ exit during quits corresponds to whenever the rat left the wait zone.

Figure 7. Neural dynamics within the OFC emerged with behavioral adaptations to a reward scarce environment (A) Time spent deciding in the offer zone before, during, and after behavioral adaptations to the reward-scarce environment. Stage 1: 1-5s condition (yellow), Stage 2: 1-10s condition (red). Stage 3: first 7 sessions of the 1-30 s reward scarce environment (light blue), reflecting the period before/during behavioral adaptations. Stage 4: final 17 sessions of the reward-scarce environment (dark blue). (B) Linger time across the four stages. (C) Probability of quitting in the wait zone (purple) and skipping in the offer zone (orange) across the four stages. (D-E) Mean (±SEM) probability of skipping in the offer zone (D) and quitting in the wait zone (E) as a function of offer delay during all four learning stages. (F) Mean (±SEM) normalized firing rate within LO at offer zone entry (orange), wait zone entry (purple), and reward delivery (brown) across the four stages. (G) Mean (±SEM) of the slope relating neural activity to time

spent in the offer zone for VO and MO across the four learning stages. (H) Mean (±SEM) second-order slopes measuring sensitivity of offer zone ramping activity to offer delay in VO and MO across the four learning stages.

Main Tables and Legends

Table 1: Summary of which subregions within the OFC exhibited neural signatures during key task states on Restaurant Row. Larger circles indicate stronger effects.

		LO	VO	MO
Reward	Anticipation	●	•	
	Delivery	●	•	
	Consumption		●	●
Offer Zone Evaluations	Entry Burst	●		
	Short-Delay Activity (Accepts)	●		
	Long-Delay Ramping Activity (Skips)		●	•
Wait Zone Reevaluations	Entry Burst	●		
	Time Remaining	●	•	
	Time Spent	●		●
	Pre-Quit Activity		●	●

Methods

Experimental Model and Study Participant Details

Subjects. Eight adult Fisher-Brown Norway rats served as subjects (4F, 4M). Because we did not obtain large cell yields from one of the rats (R665), his data was excluded from all analyses. Thus, the effective sample for the analyses presented in this paper was n = 7 (4F, 3M). Rats were single-housed in a temperature-controlled colony room with a 12 hr light/dark cycle. Each rat's daily food intake was obtained from the food earned within hour-long experimental sessions. Rats had free access to water in their home cages. All procedures were approved by the University of Minnesota Institutional Animal Care and Use Committee (IACUC).

Histology. Following completion of recording sessions, rats were euthanized via pentobarbital overdose and were then perfused. Probes were removed from the brain following perfusion. Brains were then removed and post-fixed in paraformaldehyde for 24 hours, after which they were placed in a jar containing a mixture of paraformaldehyde and sucrose for cryoprotection.

Coronal sections (40 microns) were cut using a cryostat and were subsequently stained with cresyl violet. Probe trajectory across coronal slices, coupled with records of surgical placement probe turning were used to confirm probe placements within the OFC.

Surgery. Rats were anesthetized with 2% isoflurane using an induction chamber prior to being placed on the stereotax. A custom-built head ring was attached to the skull using Metabond. The dura was removed and three craniotomies were drilled through the skull. One craniotomy was drilled directly above the medial orbital area (LO; A/P = 4.2, M/L = 0.5). Another craniotomy was drilled directly above the ventral orbital area (VO; A/P = 4.2, M/L = 1.5). The final craniotomy was drilled directly above the lateral orbital area (MO; A/P = 4.2, M/L = 2.5). Craniotomy placements were determined using the Paxinos and Watson atlas. Hemisphere placement for each OFC subregion was counterbalanced across subjects. For each rat except R581, three 32-channel Cambridge Neurotech E-2 silicon probes, each containing two shanks (9mm long with 250 microns of spacing between them), were then lowered into the cortex, one probe per craniotomy. For R581, two 64-channel Cambridge Neurotech F silicon probes were used, with 6 shanks per probe (6mm long with 200 micron spacing between each). Probes were connected to custom-built hyperdrives to allow for turning and localization. Each probe was lowered 1-1.5 mm ventral during surgery. Once probes were in place, craniotomies were filled with bone wax. The microdrives were then attached to the skull using Metabond. Finally, a custom-made skull ring and headcap were attached to the head ring to protect the implant. Following surgery, probes were turned down in small increments each day until reaching their terminal location. Most turning took place prior to recording sessions and during the first 10 recording sessions.

Behavior Training. Prior to behavioral sessions on Restaurant Row (RRow), rats were given 1 hour of access to one of the four flavored pellets (plain, chocolate, cherry, banana; 45 mg per pellet) each day until they readily consumed each flavor. Once rats were familiarized with each flavor, they were trained to run clockwise around the experimental track to earn a reward (2 flavored food pellets) at each restaurant during a 60 min experimental session (conducted once per day). During the early phase of pre-training, rewards were delivered 1 s after entering the restaurant and this delay range increased across pre-training. Specifically, once rats reliably ran the maze to earn rewards, the delay range increased to 1-5 s. Following 7 sessions of training with 1-5 s, the delay range increased to 1-10 s, which was in effect for another 7 sessions. The purpose of this training regime was to ensure that rats were approaching each restaurant, in turn, and earning enough rewards to fulfill their daily food requirements. Note that pre-training stopped at a delay range of 1-10 s.

Behavioral Recording Sessions. To ensure rats were comfortable running on the maze while plugged in to the recording equipment, rats ran up to five sessions of Restaurant Row with a 1 s delay to reward. Following this brief retraining period, experimental recording sessions began. As in training, each session lasted 60 min. In addition, rats were placed on a flowerpot for a 5 min period before and after experimental sessions to collect neural data during rest. There were a total of 35 recording sessions for each rat except R621 (23 sessions) and R626 (24 sessions), both of which ended early due to infections. The first 5 sessions consisted of the 1-5 s delay

range. The second 5 sessions (i.e., sessions 6-10) consisted of the 1-10 s delay range. The final 25 sessions consisted of the 1-30 s delay range. Previous work using Restaurant Row has shown that the 1-30 s delay range induces a host of behavioral adaptations that arise because it is no longer possible to maintain the needed daily food intake by accepting and waiting out every offer. As such, this stage of the task is referred to as the "economic crisis" and reflects a reward-scarce environment in which rats must decide how to allocate their limited time to earn their needed daily food intake. Importantly, this was the first time rats experienced such long delays in Restaurant Row.

Method Details

Neural Data Acquisition and Processing. Neural data were collected using a 256 channel Intan RHD 2000 system. Each rat was plugged into the Intan system via two headstages; a 64 channel headstage that was connected to the two-probe microdrive and a 32 channel headstage plugged into the single-probe microdrive. Voltage signals at each channel were sampled at 30 kHz.

Kilosort 2.0 was used to identify single-unit clusters within each recording session (including time on the flowerpot before and after the behavioral session). Voltage signals from each probe were normalized via median subtraction to eliminate noise and went through a 600 Hz high-pass filter to isolate potential spikes. Finally, sorting results for each probe and each session were manually curated using Phy. Each probe was analyzed separately to identify clusters.

Quantification and Statistical Analysis

The primary statistical technique used for analyses was a mixed effects ANOVA. This method was used because it provides a flexible and robust way to explore multidimensional main effects, along with interactions of hierarchically nested data and potentially correlated predictor variables[107]. All mixed effects ANOVAs were performed on data averaged across an entire session and concatenated across sessions. For this reason, we used rat identification as a random effect for all statistical tests using mixed effects models. The fixed effects depended on the specific analyses (specified below) and were most often related to either subregion (categorical; LO, VO, MO), decision type (categorical; accept vs. skip, stay vs. quit), offer delay (continuous), or learning stage (categorical; stage 1-4). We also used One-Way ANOVAs to assess simple main effects. Finally, Mann-Whitney U tests and Wilcoxon rank sum tests were used to assess whether distributions of behavioral or neural quantities were significantly different from each other or whether they were significantly different from a zero-median distribution, respectively. Specification of each statistical methods application to a given analysis is specified below.

Figure 1 Behavior Analyses: During each session, we measured the probability of skipping an offer as a function of reward delay. To do this, we found each encounter with a given delay and divided it by the number of those encounters that were skipped. This metric was collapsed across restaurants within each session. A One-Way ANOVA was used to statistically assess the effect of delay on skip probability. In addition, we assessed time spent making decisions in the offer zone by measuring, for each lap, the difference between the time at which the offer zone was exited and the time at which the offer zone was entered. We analyzed time spent in the offer zone as a

function of reward delay and decision type (i.e., accept vs. skip reaction times). A Mann-Whitney U Test was used to compare distributions of time spent making accept versus skip decisions. To statistically test decision X delay interactions, time spent for each session was normalized relative to the average time spent in the offer zone for that session. This was done to normalize between-rat differences in overall, average time spent in the offer zone. As with decision probability, all time spent data were collapsed across restaurants. A Two-Way ANOVA, with decisions and delays as factors, was then used to assess the interactive effects of time spent making each decision given offer delay.

To examine wait zone decisions, we measured for each session the probability of quitting an offer as a function of reward delay by dividing the total number of wait zone entries at a given delay by the number of encounters quit at that delay. A One-Way ANOVA was used to statistically assess the effect of delay on quit probability. We also measured the time spent to make a quit decision. This was done by taking the difference between the time at which the wait zone was exited for quit encounters only and the time the wait zone was entered during that encounter. As with offer zone behavioral metrics, all wait zone behavioral metrics were collapsed across restaurants.

Linger time was defined by the amount of time rats spent at the current restaurant after receiving reward and prior to heading to the next restaurant. This metric was calculated for each lap by taking the time at which the restaurant was exited from the time at which reward was delivered.

Figure 2 Reward Activity Analyses: We normalized each cell's firing rate by subtracting its firing rate in each temporal bin by the cell's session-wide mean firing rate and dividing by that cell's session-wide standard deviation in firing rate. A 250 ms bin size was used for calculating firing rates. All peri-event and event-related measures of neural activity used this z-scored firing rate.

Normalized firing rate was averaged across cells within each subregion during the epoch surrounding reward delivery. We explored neural dynamics during the -3 to +3 s window around the time at which the computer turned on the feeder (t = 0 s refers to the moment the feeder clicked, indicating its operation). A 250 ms bin size was used to construct peri-event activity profiles during the +/- 3 s window. The "reward anticipation" period was defined as the -3-0 s period preceding the feeder click. It took approximately 1 s for food pellets to reach the receptacle on the maze where they were available for consumption. We therefore defined the 0-1 epoch following feeder click as the "reward delivery" period, referring to the collection of sounds that happened with incoming pellets. Given how quickly rats consumed food upon its arrival into the feeding receptacle, we defined the 2-3 s epoch following feeder click as the "reward consumption" period. A Mixed Effects ANOVA, with OFC subregion and reward epoch as fixed effects and rat as a random effect, was used to statistically compare average firing rate across subregions and reward epochs.

To examine how neural activity during the reward epoch was influenced by delay, we separated each reward delivery by the delay preceding its delivery. For each cell, we found its average normalized firing rate separately within reward anticipation, delivery, and consumption

epochs for each delay. This resulted in 3 separate nCells×delays matrices, one for each reward epoch. We then averaged across cells within each subregion to obtain tuning curves that relate normalized firing rate to delay during reward anticipation, delivery, and consumption. A Mixed Effects ANOVA, with reward delay and epoch as a fixed effects and rat as a random effect, was used to examine statistically significant effects.

To explore representations of reward-specific attributes, we used a Bayesian decoding algorithm to estimate a posterior probability distribution over reward types during the reward epoch. This decoding algorithm takes as input a cells × session-time firing rate matrix along with a reward-type × cells matrix of tuning curves. Given the ensemble tuning curves, the algorithm uses ensemble firing rates at each moment, t, to estimate a probability distribution over reward types (flavors) at that moment. We constructed tuning curves to reward types during the 0-2 s epoch following reward delivery, consistent with previous research exploring reward representations in Restaurant Row[63]. We cross-validated all decoding analyses by randomly selecting half of the encounters within a session to construct tuning curves (i.e., train the model) and analyzed decoded probability distributions from the other half of the session (i.e., test the model). We repeated this cross-validation 50 times per session and averaged across them. The bin size for constructing the cells × session time firing rate matrix, along with the peri-event bin size, was 250 ms. This resulted in a sessions × time × reward-type matrix of posterior probability distributions for each subregion within the OFC.

To analyze reward-specific representations, we rotated the maze so that the posterior probability distribution over reward types were aligned to the current, next, past, and other restaurant relative to the rat's current position. For example, if the rat was currently at the restaurant offering cherry pellets, then the "current" reward would be assigned the probability of decoding to cherry. The "next" reward would be assigned the probability of decoding to banana. The "other" reward would be assigned the probability of decoding to plain. The "past" reward would be assigned the probability of decoding to chocolate. For the next encounter, the current-next-past-other assignment would be shifted 90 degrees on the maze, such that now the "current" reward was assigned the probability of decoding to banana. Thus, reward-specific representations emerged as a greater probability of decoding the current reward than the others.

To measure the strength of current reward representations during reward anticipation, delivery, and consumption, we averaged the posterior probability of each reward type across time separately within each epoch and found the difference between the probability assigned to current reward from the probability assigned to each of the other three. Thus, for each lap we obtained 3 values measuring the difference between current reward probability and next, other, and past reward probabilities. Values greater than 0 indicate that the current reward was more strongly represented than the other three. These values were averaged across each pairwise comparison and finally across laps to obtain, for each subregion within each session, a measure of the strength of reward-specific representations during each epoch surrounding reward.

As a control, we randomly assigned the rat's current restaurant with one of the three remaining labels (i.e., next, other, past). The order (or structure) of the relation between current-next-past-other was preserved for each random assignment, which resulted in a more stringent

control test because information about the order of reward flavors was preserved. In other words, the shuffle related to where within a fixed sequence of reward identities the rat was assumed to be at, rather than a complete shuffling of reward types and their spatial positions relative to the rat. As with the experimental data, we then averaged laps and across pairwise comparisons of the difference between the probability of the reward labeled as "current" with the other three reward positions. Significant current-reward encoding was determined using 1) Wilcoxon Signed Rank Tests on normalized distributions of current-reward scores, and 2) Mann-Whitney U Tests between real and shuffled distributions of current-reward scores.

Figures 3-4: Offer Zone Evaluations. We measured average normalized firing rate across cells, as described in *Figure 2: Reward* at offer zone entry and exit times to explore neural dynamics during initial evaluations. A window of -0.5 to 2.5 s around offer zone entry time was used while a window of -2.5 to + 0.5 s was used around offer zone exit time. These parameters were determined based on observations of typical occupancy time in the offer zone across rats. A Mixed Effects ANOVA, with region and time as fixed effects and rat as a random effect, was used to statistically compare neural dynamics at both zone entry and exit times.

To measure how time-dependent neural activity changed as a function of reward delay in the offer zone, we assessed average normalized firing rate broken up by offer delay for each subregion at zone entry and exit. This resulted in three separate cells × time × delay matrices, one for each subregion. We averaged across cells within each subregion to depict heatmaps of mean normalized activity throughout the offer zone across delays.

To quantify overall sensitivity of ensemble activity to reward delay, we constructed tuning curves of normalized firing rate to reward delay for each cell. Because neural dynamics observed from peri-event averages revealed time-dependent sensitivity to delay, tuning curves to reward delay were constructed separately during initial offer zone onset, to measure delay sensitivity upon immediately encountering an offer, and during the putative decision time, to measure delay sensitivity during ongoing decision processes. Tuning curves to delay during initial offer zone onset were constructed using data only from the first 1 s of time spent in the offer zone (i.e., 0-1 s after ofer zone entry). Tuning curves during the putative decision period were constructed from data excluding the first 1 s of time spent in the offer zone. This resulted in two separate cells × delay matrix of normalized firing rates, one for each epoch in the offer zone. A Mixed Effects ANOVA, with reward delay as a fixed effect and rat as a random effect, was used to statistically examine the effect of delay on tuning curves averaged across cells within each session for each subregion.

Ramping activity in the offer zone was quantified by measuring the slope relating average normalized firing rate to time spent in the offer zone. To this end, we first determined average normalized firing rate for each cell broken up by reward delay and time spent in the offer zone, resulting in a cells × delays × time-in-offer-zone matrix. We then fit a line to normalized firing rate for each cell at each delay × time-in-offer-zone combination. Slopes were subsequently averaged across time spent in the offer zone to yield a cells × delays matrix which quantified how steeply neural activity rose with time in the offer zone across delays. A Mixed Effects ANOVA, with reward delay as a fixed effect and rat as a random effect, was used to statistically

examine the effect of delay on slopes (acceleration of neural activity) relating ensemble activity to delay within each subregion. A fixed effect of region was added to the Mixed Effects model for comparisons between VO and MO.

Sensitivity of neural activity to tone presentation was determined by finding the moment corresponding to tone presentation during each offer zone encounter. Mean normalized firing rate was then measured during the -1 to 1 s epoch surrounding tone presentation across all encounters.

To examine decision-dependent neural dynamics in LO and VO throughout the offer zone, we obtained peri-event averages of normalized firing rate at offer zone entry and exit separately for accept and skip decisions. Temporal windows with respect to enter and exit times were -0.5 to 2.5 s and -2.5 to 0.5 s, respectively. Mann-Whitney U tests on distributions of normalized firing rate averaged across temporal windows for zone enter and exit times were used to determine differences in neural activity between LO and VO during accepts and skips. To further clarify how the relative activity of LO and VO changed throughout the offer zone based on which decision would be made, we measured the difference in ensemble activity between LO and VO during accept and skip encounters within each session. Specifically, we obtained an encounters × time × decisions matrix of cell-averaged normalized firing rates for each session and each subregion. We then subtracted this matrix for LO ensembles from that for VO ensembles and averaged this matrix across encounters to obtain a sessions × time × decisions matrix of LO minus VO normalized firing rates.

Figure 5-6: Wait Zone Representations and Re-evaluation. To examine sensitivity to time-remaining and time-spent for reward, we constructed 2-dimensional (time remaining × time spent) tuning curves of normalized firing rate for each cell. This resulted in 3 separate cells × time-remaining × time-spent matrices, one for each subregion. Each matrix was averaged across cells, and each path for a time-remaining × time-spent trajectory was plotted separately for the initial reward delay associated with that path to generate the triangles in Figure 5. A Mixed Effects ANOVA with time-remaining and time-spent as fixed effects and rat as a random effect was used statistically analyze tuning curves. For between-subregion comparisons, a fixed effect of region was added to the Mixed Effects model.

Decision-dependent neural dynamics within each subregion was analyzed using peri-event averages of normalized firing rate across cells at wait zone entry and exit separately based on whether the encounter was earned or quit. To this end, we constructed a cells × time × decisions matrix of normalized firing rate for each subregion. As in the offer zone, the temporal window surrounding wait zone entry was -0.5 to 2.5 s and was -2.5 to 0.5 s around wait zone exit.

The source of elevated LO activity at wait zone entry during quits was examined by comparing average normalized activity during the epoch surrounding wait zone entry (-0.5 to 0.5 s at wait zone entry) across quit, skip and earn decisions when the offer delay was long. To reduce noise resulting from sparse sampling of quit decisions that were further reduced by separating data by delay, we averaged across 11-20 s delays to define “medium delays” and 21-

30 s delays to define “long delays” for each decision. This resulted in a cells × decisions (quit, skip, earn) matrix of normalized firing rates for medium and long delays. A One-Way ANOVA was used to assess statistical differences across decisions.

Figure 7: Experience-Dependent Changes in Behavior and OFC. We divided recording sessions into four stages to explore how behavioral and neural measures adapted to the reward-scarce environment (i.e., economic crisis). The first stage consisted of the first 5 sessions of recording with a 1-5 s delay range. The second stage consisted of the next 5 sessions with a 1-10 s delay range. The third stage was the first 7 sessions of the 1-30 s delay range while the fourth stage consisted of the final 18 recording sessions with a 1-30 s delay range. The third stage reflected early periods of the reward-scarce environment in which rats adapted their behavior to the new conditions whereas the fourth stage reflected the later periods in which behavior had adapted and mostly stabilized. Of primary interest to measure behavioral adaptations to the reward scarce environment was 1) decision time in the offer zone, 2) linger time after reward delivery, 3) the number of skips in the offer zone and its dependence on delay, and 4) the number of quits in the wait zone and its dependence on delay. For each behavioral metric we averaged across all sessions within each stage. Decision time was measured within each session as the average time spent in the offer zone across laps. Linger time was measured as the average time spent in the linger zone following reward delivery before leaving the restaurant. To measure delay-dependent skips and quits, within each session we divided the number of skip/quit decisions at a given delay by the total number of offer/wait zone encounters at that delay.

Cell-averaged bursts of activity in the LO at offer, wait, and linger zone entries were measured for each session, as described in section *Figure 7: Distinct Neural Dynamics in LO and VO During Offer and Wait Zone Decisions.* We then averaged activity LO activity bursts at each zone entry across sessions within each stage to explore how these bursts changed with experience in the reward scarce.

To examine the development of offer zone ramping activity in VO and MO ensembles, we measured the slope relating ensemble activity within each subregion to time spent in the offer zone for each session, as described in *Figures 3-4: Offer Zone Evaluations.* However, we averaged slopes obtained across offer delays because we did not find any delay-dependent effects on this measure in ensembles. We then averaged neural activity slopes across sessions within each stage. A Mixed Effects ANOVA with stage as a fixed effect and rat as a random effect was used to assess significant main effects. Comparisons between VO and MO similarly used a Mixed Effects ANOVA but with subregion included as an additional fixed effect.

The emergence of delay-dependent ramping in VO and MO across the reward-scarce environment was measured by finding the slope relating ensemble activity within each subregion to offer delay for each session. We then fit a line to obtain a second-order slope that measured the degree to which VO ramping activity was sensitive to delay. For example, a second-order slope of 0 would mean that VO ramping activity was unrelated to delay, whereas a slope > 0 means that VO activity accelerated more steeply for longer delays. We then averaged this second order slope across sessions within each stage. A Mixed Effects ANOVA with stage as a fixed effect and ratID as a random effect was used to assess significant main effects. Comparisons between VO

and MO similarly used a Mixed Effects ANOVA but with subregion included as an additional fixed effect.

Figure S5: Distinct Neural Dynamics in LO and VO During Offer and Wait Zone Decisions. Each cell's average normalized firing rate during an epoch of interest was used to determine the degree to which that cell was sensitive to the behavioral event defining that epoch. For LO cells, we were interested in sensitivity to offer zone entry, wait zone entry, and reward delivery (linger zone entry). Normalized firing rate for each LO cell was averaged across a -0.5 to 0.5 s window at offer zone entry, a 0.5 to 1 s window at wait zone entry, and a 0-1 s window at reward delivery. For VO cells, we were interested in offer zone exit given skip, wait zone exit given quit, and linger zone exit. Normalized firing rate for each VO cell was averaged during the -2.5 to 0.5 s window prior leaving each of the three zones. Temporal windows used for averaging neural activity were determined based on ensemble-level neural dynamics at each behavioral epoch. A cell was deemed sensitive to an event if its normalized firing rate exceeded 1 standard deviation of the distribution mean of cell scores at that event. This method thereby selected those cells within the population that were most strongly active during the behavioral epoch of interest. The correlation of normalized firing rate scores between each pair of behavioral epochs (offer zone entry, wait zone entry, and reward) was measured to determine the degree to which a cell's activity at one epoch was related to its activity at another. In addition, we identified cells that were deemed sensitive to a behavioral epoch and plotted peri-event averages of activity within those cells at each behavioral epoch. In doing so, we were able to compare how the temporal dynamics of a collection of cells deemed sensitive to an event unfolded across behavioral epochs.

To examine the extent to which each cell's tuning curve to momentary delay was consistent between the offer zone and wait zone, we normalized the correlation between tuning curves derived from each zone relative to a shuffled distribution. Specifically, for each cell, we preserved the relation between firing rate and momentary delay for a given zone (e.g., offer zone) and then randomly shuffled the relation between firing rate and momentary delay for the other zone (e.g., wait zone). We then found the correlation between these tuning curves. This process was repeated 100 times. The true correlation between tuning curves was then z-scored relative to the n = 100 shuffled distributions.

References

1. Charnov, E. L. Optimal foraging, the marginal value theorem. *Theor. Popul. Biol.* **9**, 129 136 (1976).

2. Stephens, D. W. Decision ecology: Foraging and the ecology of animal decision making. *Cogn. Affect. Behav. Neurosci.* **8**, 475–484 (2008).

3. Stephens, D. W. & Krebs, J. R. *Foraging Theory*. (1986).

4. Calhoun, A. J. & Hayden, B. Y. The foraging brain. *Curr. Opin. Behav. Sci.* **5**, 24–31 (2015).

5. Kacelnik, A., Vasconcelos, M., Monteiro, T. & Aw, J. Darwin's "tug-of-war" vs. starlings' "horse-racing": how adaptations for sequential encounters drive simultaneous choice. *Behav. Ecol. Sociobiol.* **65**, 547–558 (2011).

6. Mobbs, D., Trimmer, P. C., Blumstein, D. T. & Dayan, P. Foraging for foundations in decision neuroscience: insights from ethology. *Nat. Rev. Neurosci.* **19**, 419–427 (2018).

7. Pearson, J. M., Watson, K. K. & Platt, M. L. Decision Making: The Neuroethological Turn. *Neuron* **82**, 950–965 (2014).

8. Pyke, G. Optimal Foraging Theory: A Critical Review. *Annu. Rev. Ecol. Syst.* **15**, 523 575 (1984).

9. Fellows, L. K. Orbitofrontal contributions to value-based decision making: evidence from humans with frontal lobe damage. *Ann. N. Y. Acad. Sci.* **1239**, 51–58 (2011).

10. Padoa-Schioppa, C. & Conen, K. E. Orbitofrontal Cortex: A Neural Circuit for Economic Decisions. *Neuron* **96**, 736–754 (2017).

11. Rudebeck, P. H. & Izquierdo, A. Foraging with the frontal cortex: A cross-species evaluation of reward-guided behavior. *Neuropsychopharmacology* **47**, 134–146 (2022).

12. Wallis, J. D. Orbitofrontal Cortex and Its Contribution to Decision-Making. *Annu. Rev. Neurosci.* **30**, 31–56 (2007).

13. Wilson, R. C., Takahashi, Y. K., Schoenbaum, G. & Niv, Y. Orbitofrontal Cortex as a Cognitive Map of Task Space. *Neuron* **81**, 267–279 (2014).

14. Elliott, R. Dissociable Functions in the Medial and Lateral Orbitofrontal Cortex: Evidence from Human Neuroimaging Studies. *Cereb. Cortex* **10**, 308–317 (2000).

15. Izquierdo, A. Functional Heterogeneity within Rat Orbitofrontal Cortex in Reward Learning and Decision Making. *J. Neurosci.* **37**, 10529–10540 (2017).

16. Rich, E. L. & Wallis, J. D. Medial-lateral Organization of the Orbitofrontal Cortex. *J. Cogn. Neurosci.* **26**, 1347–1362 (2014).

17. Rudebeck, P. H. & Murray, E. A. Dissociable Effects of Subtotal Lesions within the Macaque Orbital Prefrontal Cortex on Reward-Guided Behavior. *J. Neurosci.* **31**, 10569 10578 (2011).

18. Zald, D. H. *et al.* Meta-Analytic Connectivity Modeling Reveals Differential Functional Connectivity of the Medial and Lateral Orbitofrontal Cortex. *Cereb. Cortex* **24**, 232–248 (2014).

19. Barreiros, I. V., Panayi, M. C. & Walton, M. E. Organization of Afferents along the Anterior–posterior and Medial–lateral Axes of the Rat Orbitofrontal Cortex. *Neuroscience* **460**, 53–68 (2021).

20. Heilbronner, S. R., Rodriguez-Romaguera, J., Quirk, G. J., Groenewegen, H. J. & Haber, S. N. Circuit-Based Corticostriatal Homologies Between Rat and Primate. *Biol. Psychiatry* **80**, 509–521 (2016).

21. Hoover, W. B. & Vertes, R. P. Projections of the medial orbital and ventral orbital cortex in the rat. *J. Comp. Neurol.* **519**, 3766–3801 (2011).

22. Murphy, M. J. M. & Deutch, A. Y. Organization of afferents to the orbitofrontal cortex in the rat. *J. Comp. Neurol.* **526**, 1498–1526 (2018).

23. Banerjee, A. *et al.* Value-guided remapping of sensory cortex by lateral orbitofrontal cortex. *Nature* **585**, 245–250 (2020).

24. Gremel, C. M. & Costa, R. M. Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. *Nat. Commun.* **4**, 2264 (2013).

25. Hong, D.-D. *et al.* Neurons in rat orbitofrontal cortex and medial prefrontal cortex exhibit distinct responses in reward and strategy-update in a risk-based decision-making task. *Metab. Brain Dis.* **34**, 417–429 (2019).

26. Kepecs, A., Uchida, N., Zariwala, H. A. & Mainen, Z. F. Neural correlates, computation and behavioural impact of decision confidence. *Nature* **455**, 227–231 (2008).

27. Lak, A. *et al.* Orbitofrontal Cortex Is Required for Optimal Waiting Based on Decision Confidence. *Neuron* **84**, 190–201 (2014).

28. Sadacca, B. F. *et al.* Orbitofrontal neurons signal sensory associations underlying model based inference in a sensory preconditioning task. *eLife* **7**, e30373 (2018).

29. Stalnaker, T. A. *et al.* Orbitofrontal neurons infer the value and identity of predicted outcomes. *Nat. Commun.* **5**, 3926 (2014).

30. Zeeb, F. D., Floresco, S. B. & Winstanley, C. A. Contributions of the orbitofrontal cortex to impulsive choice: interactions with basal levels of impulsivity, dopamine signalling, and reward-related cues. *Psychopharmacology (Berl.)* **211**, 87–98 (2010).

31. 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).

32. Bradfield, L. A., Hart, G. & Balleine, B. W. Inferring action-dependent outcome representations depends on anterior but not posterior medial orbitofrontal cortex. *Neurobiol. Learn. Mem.* **155**, 463–473 (2018).

33. Burton, A. C., Kashtelyan, V., Bryden, D. W. & Roesch, M. R. Increased Firing to Cues That Predict Low-Value Reward in the Medial Orbitofrontal Cortex. *Cereb. Cortex* **24**, 3310–3321 (2014).

34. Jenni, N. L., Rutledge, G. & Floresco, S. B. Distinct Medial Orbitofrontal–Striatal Circuits Support Dissociable Component Processes of Risk/Reward Decision-Making. *J. Neurosci.* **42**, 2743–2755 (2022).

35. Lichtenberg, N. T. *et al.* The Medial Orbitofrontal Cortex–Basolateral Amygdala Circuit Regulates the Influence of Reward Cues on Adaptive Behavior and Choice. *J. Neurosci.* **41**, 7267–7277 (2021).

36. Lopatina, N. *et al.* Medial Orbitofrontal Neurons Preferentially Signal Cues Predicting Changes in Reward during Unblocking. *J. Neurosci.* **36**, 8416–8424 (2016).

37. Lopatina, N. *et al.* Ensembles in medial and lateral orbitofrontal cortex construct cognitive maps emphasizing different features of the behavioral landscape. *Behav. Neurosci.* **131**, 201–212 (2017).

38. Malvaez, M., Shieh, C., Murphy, M. D., Greenfield, V. Y. & Wassum, K. M. Distinct cortical–amygdala projections drive reward value encoding and retrieval. *Nat. Neurosci.* **22**, 762–769 (2019).

39. Bradfield, L. A. & Hart, G. Rodent medial and lateral orbitofrontal cortices represent unique components of cognitive maps of task space. *Neurosci. Biobehav. Rev.* **108**, 287 294 (2020).

40. Hervig, M. E. *et al.* Dissociable and Paradoxical Roles of Rat Medial and Lateral Orbitofrontal Cortex in Visual Serial Reversal Learning. *Cereb. Cortex* **30**, 1016–1029 (2020).

41. Noonan, M. P. *et al.* Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. *Proc. Natl. Acad. Sci.* **107**, 20547–20552 (2010).

42. Noonan, M. P., Chau, B. K. H., Rushworth, M. F. S. & Fellows, L. K. Contrasting Effects of Medial and Lateral Orbitofrontal Cortex Lesions on Credit Assignment and Decision Making in Humans. *J. Neurosci.* **37**, 7023–7035 (2017).

43. Brockett, A. T., Kumar, N., Sharalla, P. & Roesch, M. R. Optogenetic Inhibition of the Orbitofrontal Cortex Disrupts Inhibitory Control during Stop-Change Performance in Male Rats. *eneuro* **11**, ENEURO.0015-24.2024 (2024).

44. Hardung, S. *et al.* A Functional Gradient in the Rodent Prefrontal Cortex Supports Behavioral Inhibition. *Curr. Biol.* **27**, 549–555 (2017).

45. Kheramin, S. *et al.* Role of the orbital prefrontal cortex in choice between delayed and uncertain reinforcers: a quantitative analysis. *Behav. Processes* **64**, 239–250 (2003).

46. Kheramin, S. *et al.* Effects of orbital prefrontal cortex dopamine depletion on inter temporal choice: a quantitative analysis. *Psychopharmacology (Berl.)* **175**, (2004).

47. Meyer, H. C. & Bucci, D. J. Imbalanced Activity in the Orbitofrontal Cortex and Nucleus Accumbens Impairs Behavioral Inhibition. *Curr. Biol.* **26**, 2834–2839 (2016).

48. S., K. *et al.* Effects of quinolinic acid-induced lesions of the orbital prefrontal cortex on inter-temporal choice: a quantitative analysis. *Psychopharmacology (Berl.)* **165**, 9–17 (2002).

49. Stolyarova, A. & Izquierdo, A. Complementary contributions of basolateral amygdala and orbitofrontal cortex to value learning under uncertainty. *eLife* **6**, e27483 (2017).

50. Diehl, G. W. & Redish, A. D. Differential processing of decision information in subregions of rodent medial prefrontal cortex. *eLife* **12**, e82833 (2023).

51. Redish, A. D. *et al.* Sunk cost sensitivity during change-of-mind decisions is informed by both the spent and remaining costs. *Commun. Biol.* **5**, 1337 (2022).

52. Schmidt, B., Duin, A. A. & Redish, A. D. Disrupting the medial prefrontal cortex alters hippocampal sequences during deliberative decision making. *J. Neurophysiol.* **121**, 1981 2000 (2019).

53. Schmidt, B. & Redish, A. D. Disrupting the medial prefrontal cortex with DESIGNER RECEPTORS EXCLUSIVELY ACTIVATED BY DESIGNER DRUG alters hippocampal sharp-wave ripples and their associated cognitive processes. *Hippocampus* **31**, 1051–1067 (2021).

54. Sweis, B. M. *et al.* Sensitivity to “sunk costs” in mice, rats, and humans. *Science* **361**, 178–181 (2018).

55. Sweis, B. M., Larson, E. B., Redish, A. D. & Thomas, M. J. Altering gain of the infralimbic-to-accumbens shell circuit alters economically dissociable decision-making algorithms. *Proc. Natl. Acad. Sci.* **115**, (2018).

56. Sweis, B. M., Redish, A. D. & Thomas, M. J. Prolonged abstinence from cocaine or morphine disrupts separable valuations during decision conflict. *Nat. Commun.* **9**, 2521 (2018).

57. Sweis, B. M., Thomas, M. J. & Redish, A. D. Mice learn to avoid regret. *PLOS Biol.* **16**, e2005853 (2018).

58. Schoenbaum, G., Chiba, A. A. & Gallagher, M. Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nat. Neurosci.* **1**, 155–159 (1998).

59. Schoenbaum, G., Chiba, A. A. & Gallagher, M. Neural Encoding in Orbitofrontal Cortex and Basolateral Amygdala during Olfactory Discrimination Learning. *J. Neurosci.* **19**, 1876–1884 (1999).

60. Schoenbaum, G., Chiba, A. A. & Gallagher, M. Changes in Functional Connectivity in Orbitofrontal Cortex and Basolateral Amygdala during Learning and Reversal Training. *J. Neurosci.* **20**, 5179–5189 (2000).

61. Schoenbaum, G. & Eichenbaum, H. Information coding in the rodent prefrontal cortex. I. Single-neuron activity in orbitofrontal cortex compared with that in pyriform cortex. *J. Neurophysiol.* **74**, 733–750 (1995).

62. Schoenbaum, G. & Eichenbaum, H. Information coding in the rodent prefrontal cortex. II. Ensemble activity in orbitofrontal cortex. *J. Neurophysiol.* **74**, 751–762 (1995).

63. Steiner, A. P. & Redish, A. D. Behavioral and neurophysiological correlates of regret in rat decision-making on a neuroeconomic task. *Nat. Neurosci.* **17**, 995–1002 (2014).

64. Stott, J. J. & Redish, A. D. A functional difference in information processing between orbitofrontal cortex and ventral striatum during decision-making behaviour. *Philos. Trans. R. Soc. B Biol. Sci.* **369**, 20130472 (2014).

65. Roesch, M. R., Taylor, A. R. & Schoenbaum, G. Encoding of Time-Discounted Rewards in Orbitofrontal Cortex Is Independent of Value Representation. *Neuron* **51**, 509–520 (2006).

66. Basu, R. *et al.* The orbitofrontal cortex maps future navigational goals. *Nature* **599**, 449 452 (2021).

67. Howard, J. D., Gottfried, J. A., Tobler, P. N. & Kahnt, T. Identity-specific coding of future rewards in the human orbitofrontal cortex. *Proc. Natl. Acad. Sci.* **112**, 5195–5200 (2015).

68. Howard, J. D. & Kahnt, T. Identity-Specific Reward Representations in Orbitofrontal Cortex Are Modulated by Selective Devaluation. *J. Neurosci.* **37**, 2627–2638 (2017).

69. Howard, J. D. & Kahnt, T. To be specific: The role of orbitofrontal cortex in signaling reward identity. *Behav. Neurosci.* **135**, 210–217 (2021).

70. McDannald, M. A. *et al.* Orbitofrontal neurons acquire responses to 'valueless' Pavlovian cues during unblocking. *eLife* **3**, e02653 (2014).

71. Rudebeck, P. H. & Murray, E. A. The Orbitofrontal Oracle: Cortical Mechanisms for the Prediction and Evaluation of Specific Behavioral Outcomes. *Neuron* **84**, 1143–1156 (2014).

72. Suzuki, S., Cross, L. & O'Doherty, J. P. Elucidating the underlying components of food valuation in the human orbitofrontal cortex. *Nat. Neurosci.* **20**, 1780–1786 (2017).

73. Zhang, K., Ginzburg, I., McNaughton, B. L. & Sejnowski, T. J. Interpreting Neuronal Population Activity by Reconstruction: Unified Framework With Application to Hippocampal Place Cells. *J. Neurophysiol.* **79**, 1017–1044 (1998).

74. Gardner, M. P. H. *et al.* Processing in Lateral Orbitofrontal Cortex Is Required to Estimate Subjective Preference during Initial, but Not Established, Economic Choice. *Neuron* **108**, 526-537.e4 (2020).

75. Kuwabara, M., Kang, N., Holy, T. E. & Padoa-Schioppa, C. Neural mechanisms of economic choices in mice. *eLife* **9**, e49669 (2020).

76. Levy, D. J. & Glimcher, P. W. Comparing Apples and Oranges: Using Reward-Specific and Reward-General Subjective Value Representation in the Brain. *J. Neurosci.* **31**, 14693–14707 (2011).

77. Padoa-Schioppa, C. Range-Adapting Representation of Economic Value in the Orbitofrontal Cortex. *J. Neurosci.* **29**, 14004–14014 (2009).

78. Padoa-Schioppa, C. & Assad, J. A. Neurons in the orbitofrontal cortex encode economic value. *Nature* **441**, 223–226 (2006).

79. Padoa-Schioppa, C. & Assad, J. A. Neurons in the orbitofrontal cortex encode economic value. *Nature* **441**, 223–226 (2006).

80. Rich, E. L. & Wallis, J. D. Decoding subjective decisions from orbitofrontal cortex. *Nat. Neurosci.* **19**, 973–980 (2016).

81. Green, L. & Myerson, J. A Discounting Framework for Choice With Delayed and Probabilistic Rewards. *Psychol. Bull.* **130**, 769–792 (2004).

82. Schultz, W. Dopamine signals for reward value and risk: basic and recent data. *Behav. Brain Funct.* **6**, 24 (2010).

83. Constantino, S. M. & Daw, N. D. Learning the opportunity cost of time in a patch foraging task. *Cogn. Affect. Behav. Neurosci.* **15**, 837–853 (2015).

84. Krebs, J. R., Kacelnik, A. & Taylor, P. Test of optimal sampling by foraging great tits. *Nature* **275**, 27–31 (1978).

85. Otto, A. R. & Daw, N. D. The opportunity cost of time modulates cognitive effort. *Neuropsychologia* **123**, 92–105 (2019).

86. Stephens, D. W. & Krebs, J. R. *Foraging Theory*. (Princeton University Press, 1986).

87. Namboodiri, V. M. K. *et al.* Single-cell activity tracking reveals that orbitofrontal neurons acquire and maintain a long-term memory to guide behavioral adaptation. *Nat. Neurosci.* **22**, 1110–1121 (2019).

88. Riceberg, J. S. & Shapiro, M. L. Reward Stability Determines the Contribution of Orbitofrontal Cortex to Adaptive Behavior. *J. Neurosci.* **32**, 16402–16409 (2012).

89. Riceberg, J. S. & Shapiro, M. L. Orbitofrontal Cortex Signals Expected Outcomes with Predictive Codes When Stable Contingencies Promote the Integration of Reward History. *J. Neurosci.* **37**, 2010–2021 (2017).

90. Schoenbaum, G., Nugent, S. L., Saddoris, M. P. & Setlow, B. Orbitofrontal lesions in rats impair reversal but not acquisition of go, no-go odor discriminations. *NeuroReport* **13**, 885–890 (2002).

91. Stalnaker, T. A., Raheja, N. & Schoenbaum, G. Orbitofrontal State Representations Are Related to Choice Adaptations and Reward Predictions. *J. Neurosci.* **41**, 1941–1951 (2021).

92. Young, J. J. & Shapiro, M. L. Dynamic Coding of Goal-Directed Paths by Orbital Prefrontal Cortex. *J. Neurosci.* **31**, 5989–6000 (2011).

93. Castellano, S. Putting mechanisms in foraging theory: the role of computational mechanisms in optimal decision making. *Anim. Behav.* **153**, 159–169 (2019).

94. Mah, A., Schiereck, S. S., Bossio, V. & Constantinople, C. M. Distinct value computations support rapid sequential decisions. *Nat. Commun.* **14**, 7573 (2023).

95. Werner, E. E. & Hall, D. J. Optimal Foraging and the Size Selection of Prey by the Bluegill Sunfish (Lepomis Macrochirus). *Ecology* **55**, 1042–1052 (1974).

96. Kahnt, T. & Schoenbaum, G. Cross-species studies on orbitofrontal control of inference based behavior. *Behav. Neurosci.* **135**, 109–119 (2021).

97. Wallis, J. D. Cross-species studies of orbitofrontal cortex and value-based decision making. *Nat. Neurosci.* **15**, 13–19 (2012).

98. Ongur, D. The Organization of Networks within the Orbital and Medial Prefrontal Cortex of Rats, Monkeys and Humans. *Cereb. Cortex* **10**, 206–219 (2000).

99. Schoenbaum, G., Roesch, M. R. & Stalnaker, T. A. Orbitofrontal cortex, decision-making and drug addiction. *Trends Neurosci.* **29**, 116–124 (2006).

100. Passingham, R. E. & Wise, S. P. *The Neurobiology of the Prefrontal Cotex: Anatomy, Evolution, and the Origin of Insight.* (Oxford University Press, 2012).

101. Padoa-Schioppa, C. & Conen, K. E. Orbitofrontal Cortex: A Neural Circuit for Economic Decisions. *Neuron* **96**, 736–754 (2017).

102. Zhou, J., Gardner, M. P. & Schoenbaum, G. Is the core function of orbitofrontal cortex to signal values or make predictions? *Curr. Opin. Behav. Sci.* **41**, 1–9 (2021).

103. Kringelbach, M. L. The human orbitofrontal cortex: linking reward to hedonic experience. *Nat. Rev. Neurosci.* **6**, 691–702 (2005).

104. Abram, S. V., Hanke, M., Redish, A. D. & MacDonald, A. W. Neural signatures underlying deliberation in human foraging decisions. *Cogn. Affect. Behav. Neurosci.* **19**, 1492–1508 (2019).

105. Abram, S. V., Breton, Y.-A., Schmidt, B., Redish, A. D. & MacDonald, A. W. The Web Surf Task: A translational model of human decision-making. *Cogn. Affect. Behav. Neurosci.* **16**, 37–50 (2016).

106. Kazinka, R., MacDonald, A. W. & Redish, A. D. Sensitivity to Sunk Costs Depends on Attention to the Delay. *Front. Psychol.* **12**, 604843 (2021).

107. Yu, Z. *et al.* Beyond t test and ANOVA: applications of mixed-effects models for more rigorous statistical analysis in neuroscience research. *Neuron* **110**, 21–35 (2022).

Figure 1. Rats made economic sequential foraging decisions in Restaurant Row (A) In Restaurant Row, rats make sequential decisions in the "offer zone" (OZ) and "wait zone" (WZ) based on the delay to reward. (B) Mean (±SEM) probability of skipping as a function of delay. (C) Probability density function of time spent making accept (blue) and skip (orange) decisions. (D-E) Cumulative density functions of time spent making accept (D) and skip (E) decisions across delays. Blue = shorter delays; Red = longer delays. (F) Mean (±SEM) probability of quitting as a function of delay. (G) Probability density function of time to quit.

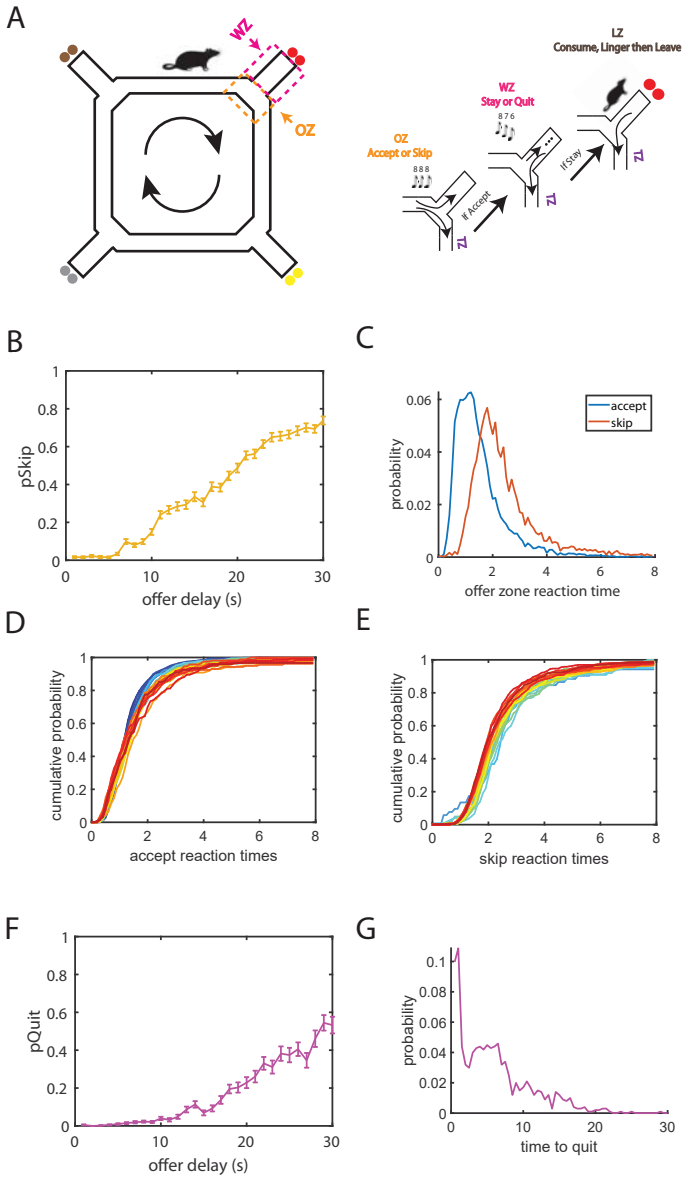

A
WZ
OZ
LZ
Consume, Linger then Leave
WZ
Stay or Quit
OZ
Accept or Skip
If Accept
If Stay
B
pSkip
offer delay (s)
C
accept
skip
probability
offer zone reaction time
D
cumulative probability
accept reaction times
E
cumulative probability
skip reaction times
F
pQuit
offer delay (s)
G
probability
time to quit

Figure 2. LO, VO, and MO encoded reward-specific information through distinct neural dynamics during reward (A) Mean (±SEM) normalized firing rate within LO, VO, and MO aligned to reward delivery. (B-D) Mean (±SEM) normalized firing rate as a function of offer delay during anticipation (B), delivery (C), and consumption (D). (E) Illustration of decoding analysis, current restaurant = location of the rat. Next, past, and other restaurants rotate with the current location for analysis. (F-H) Mean (±SEM) probability of decoding to each of the four rewards in LO (F), VO (G), and MO (H).

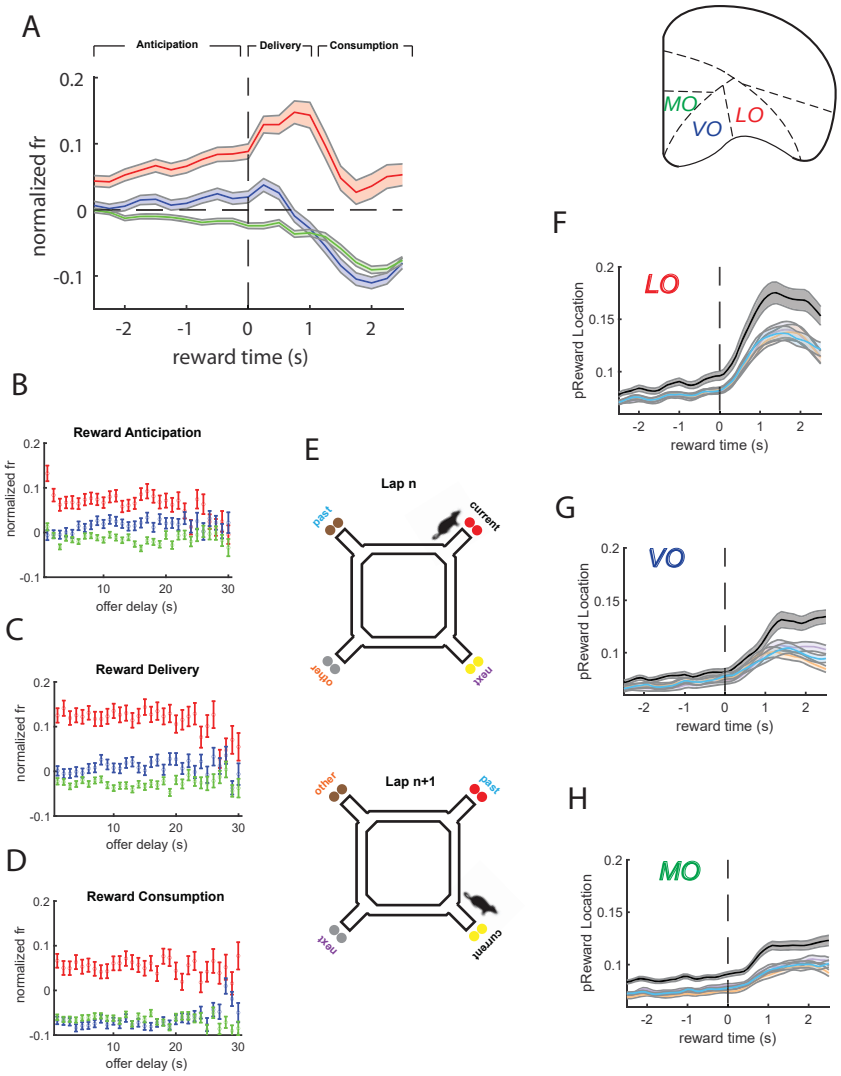

A
Anticipation
Delivery
Consumption
normalized fr
reward time (s)
B
Reward Anticipation
normalized fr
offer delay (s)
C
Reward Delivery
normalized fr
offer delay (s)
D
Reward Consumption
normalized fr
offer delay (s)
E
Lap n
past
current
other
next
Lap n+1
other
past
next
current
MO
VO
LO
F
LO
pReward Location
reward time (s)
G
VO
pReward Location
reward time (s)
H
MO
pReward Location
reward time (s)

Figure 3. Distinct neural dynamics and representational content in LO, VO, and MO during initial evaluations (A-C) Mean normalized firing for LO (A), VO (B), and MO (C) aligned to OZ entry (left) and exit (right). Top panels show mean (±SEM) activity across all encounters while bottom panels show mean activity broken up by delay. (D-E). Mean (±SEM) tuning curves of LO, VO, and MO ensembles to offer delay when entering the OZ (D) and during the putative decision period (E). (F) Mean (±SEM) slopes relating normalized firing rate to OZ time, broken up by offer delay for LO, VO, and MO. (G) Mean (± SEM) normalized firing rate aligned to tone presentation at t = 0 and t = -1s. Tone presentations lasted 0.5 s and are represented as filled bars along the top of the plot.

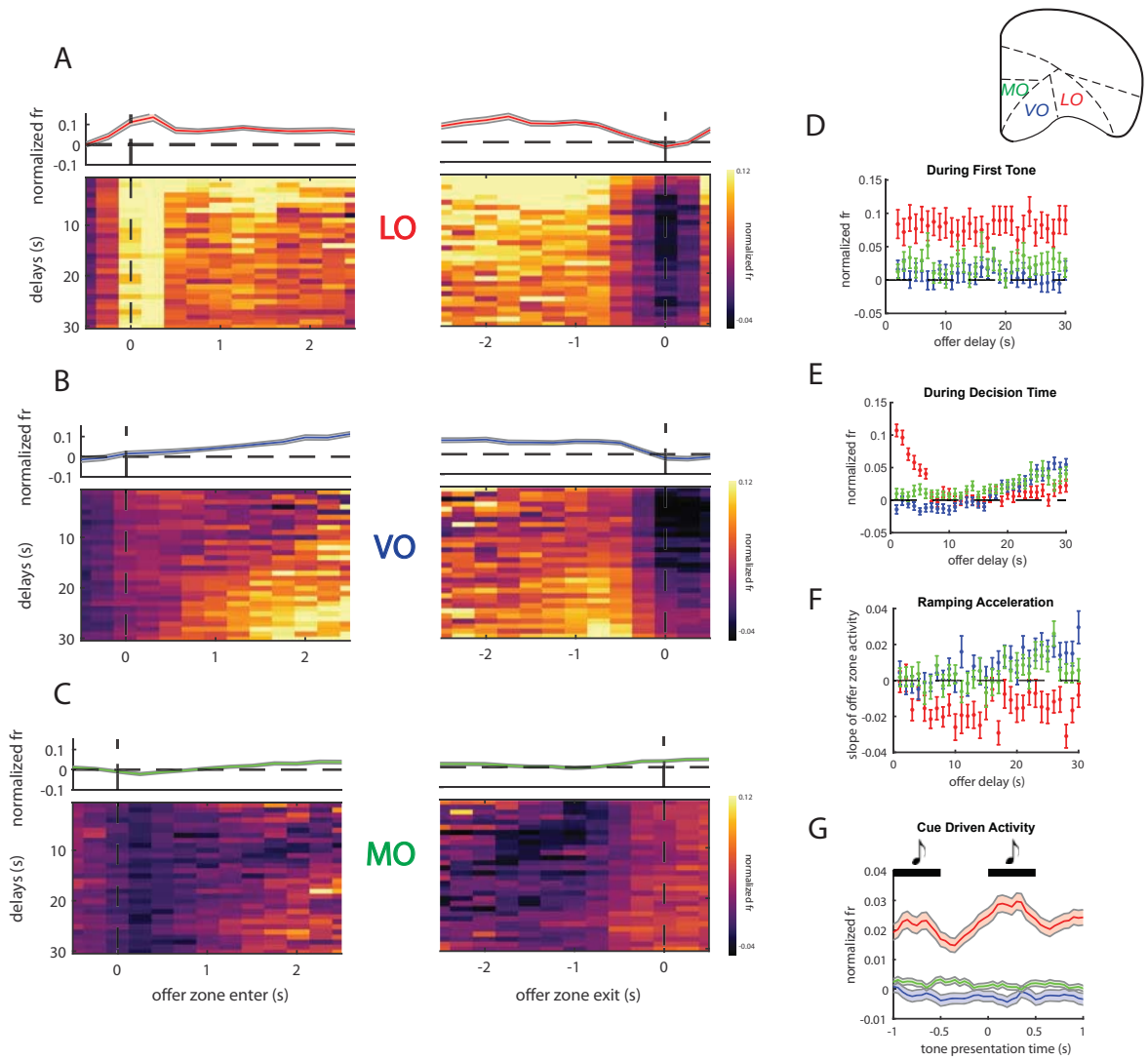

A
B
C
LO
VO
MO
normalized fr
delays (s)
offer zone enter (s)
offer zone exit (s)
D
E
F
G
MO
VO
LO
During First Tone
During Decision Time
Ramping Acceleration
Cue Driven Activity
offer delay (s)
slope of offer zone activity
tone presentation time (s)

Figure 4. Opposing neural dynamics in LO and VO reflected participation in distinct systems during initial evaluations (A-B) Mean (± SEM) normalized firing rate in LO and VO during accepted encounters aligned to OZ entry (A) and exit. (B). (C-D) Mean (±SEM) normalized firing rate in LO and VO during skip encounters aligned to OZ entry (C) and exit (D). (E-F) Mean (±SEM) difference in normalized firing rate between LO and VO at each moment in time during accept and skip encounters aligned to OZ entry (E) and exit (F).

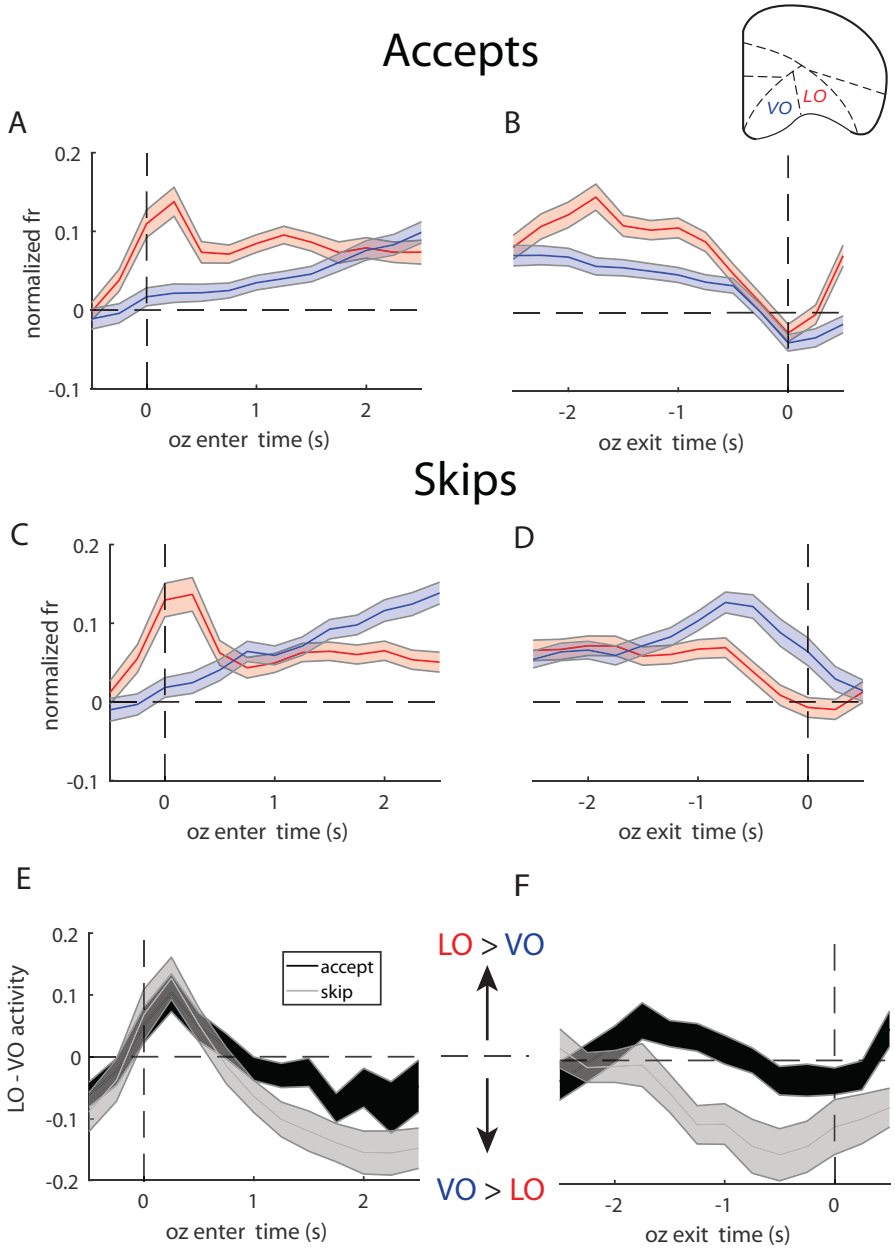

Accepts
VO
LO
A
B
normalized fr
0.2
0.1
0
-0.1
0
1
2
oz enter time (s)
-2
-1
0
oz exit time (s)
Skips
C
D
E
F
LO - VO activity
-0.2
accept
skip
LO > VO
VO > LO

Figure 5. LO, VO, and MO differentially encoded time remaining and time spent during the wait zone (A-C) Mean tuning curves to time remaining and time spent for LO (A), VO (B), and MO (C). Each pixel in the 3-dimensional space reflects mean firing rate defined by time remaining (z-axis) and time spent (x-axis) given the initial offer delay (y-axis) at that moment. Moving from top to bottom along with z-axis means time to reward is approaching. Moving from back to front on the x-axis means more time has already been spent waiting for reward. (D-E). Mean (±SEM) tuning curves to time remaining (D) and time spent (E) for LO, VO, and MO ensembles. (F) Mean (±SEM) normalized firing rate aligned to tone presentation at t = 0 and t = -1. Tone presentations lasted 0.5 s and are indicated by filled bars along the top of the plot.

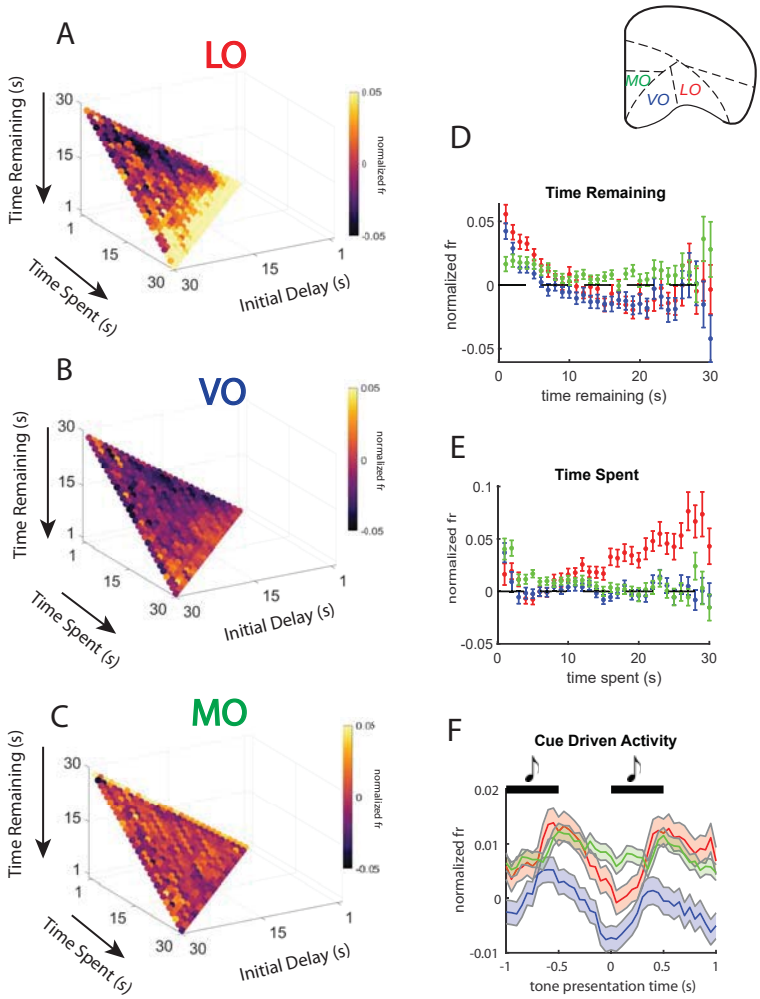

A
LO
Time Remaining (s)
Time Spent (s)
Initial Delay (s)
normalized fr
B
VO
C
MO
D
Time Remaining
normalized fr
time remaining (s)
E
Time Spent
time spent (s)
F
Cue Driven Activity
tone presentation time (s)
MO
VO
LO

Figure 6. Differential contributions of LO, VO, and MO to reevaluative change-of-mind decisions (A-B) Mean (±SEM) normalized firing rate in LO during stay (light lines) and quit (dark lines) encounters oriented with respect to WZ enter time (A) and exit time (B). (C-D) Mean (±SEM) normalized firing rate in VO during stay and quit encounters oriented with respect to WZ entry (C) and exit (D). (E-F) Mean (±SEM) normalized firing rate in MO during stay and quit encounters aligned to WZ entry (E) and exit (F). Note that WZ exit during stay encounters corresponds to reward delivery, whereas WZ exit during quits corresponds to whenever the rat left the wait zone.

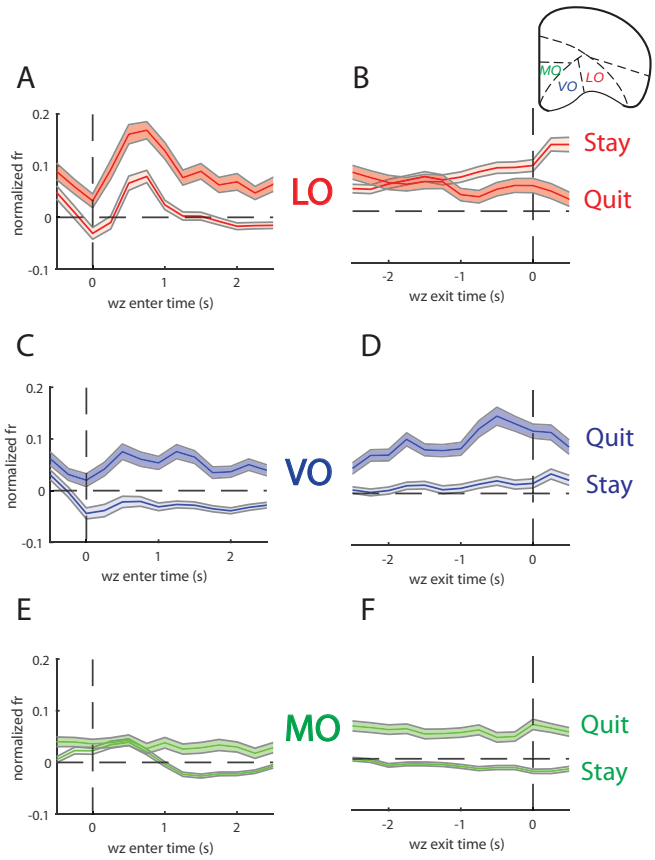
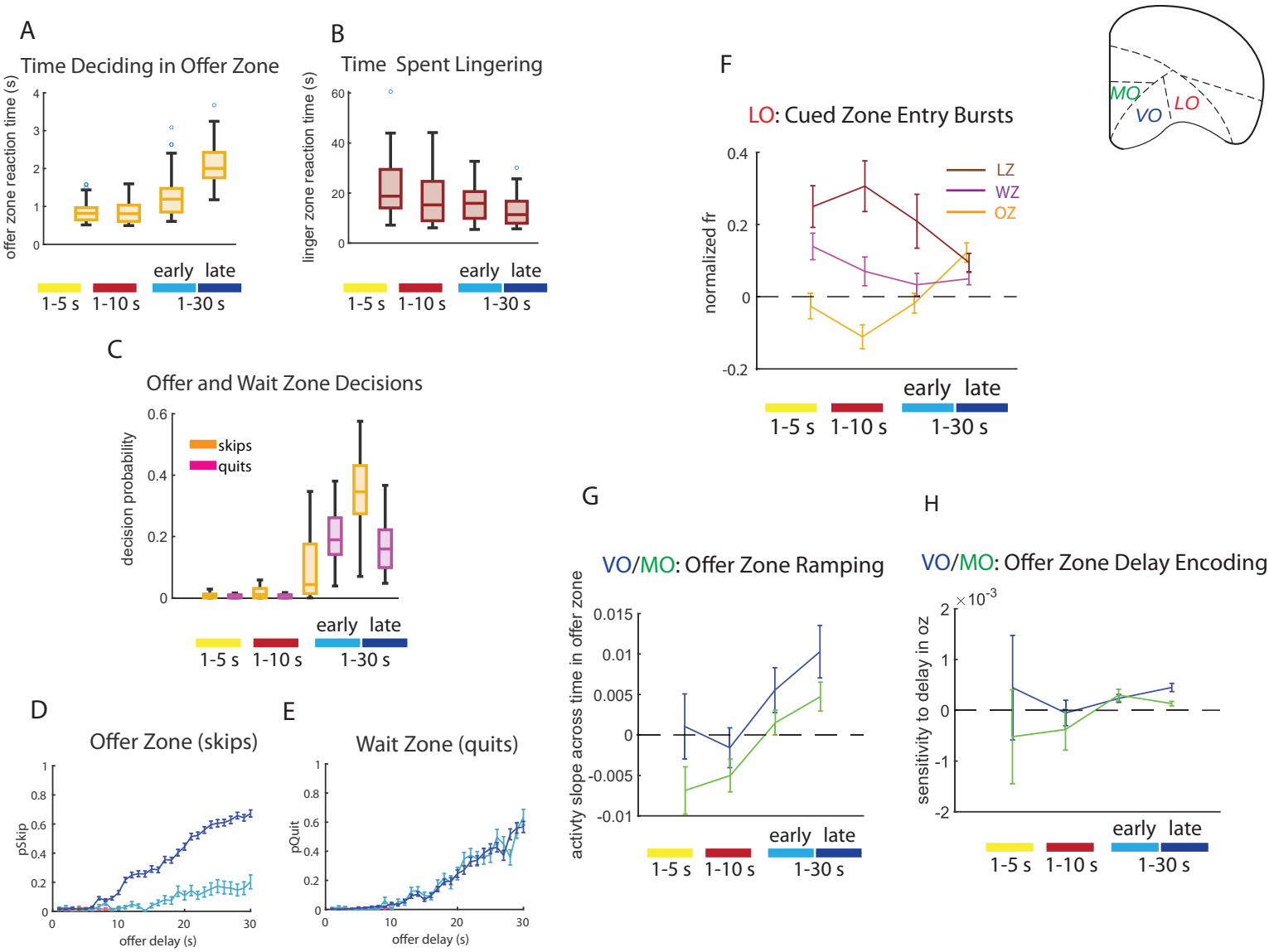

A
B
MO
VO
LO
normalized fr
0.2
0.1
0
-0.1
0
1
2
wz enter time (s)
LO
Stay
Quit
-2
-1
0
wz exit time (s)
C
D
VO
Quit
Stay
E
F
MO
Quit
Stay

Figure 7. Neural dynamics within the OFC emerged with behavioral adaptations to a reward scarce environment (A) Time spent deciding in the offer zone before, during, and after behavioral adaptations to the reward-scarce environment. Stage 1: 1-5s condition (yellow), Stage 2: 1-10s condition (red). Stage 3: first 7 sessions of the 1-30 s reward scarce environment (light blue), reflecting the period before/during behavioral adaptations. Stage 4: final 17 sessions of the reward-scarce environment (dark blue). (B) Linger time across the four stages. (C) Probability of quitting in the wait zone (purple) and skipping in the offer zone (orange) across the four stages. (D-E) Mean (±SEM) probability of skipping in the offer zone (D) and quitting in the wait zone (E) as a function of offer delay during all four learning stages. (F) Mean (±SEM) normalized firing rate within LO at offer zone entry (orange), wait zone entry (purple), and reward delivery (brown) across the four stages. (G) Mean (±SEM) of the slope relating neural activity to time spent in the offer zone for VO and MO across the four learning stages. (H) Mean (±SEM) second-order slopes measuring sensitivity of offer zone ramping activity to offer delay in VO and MO across the four learning stages.

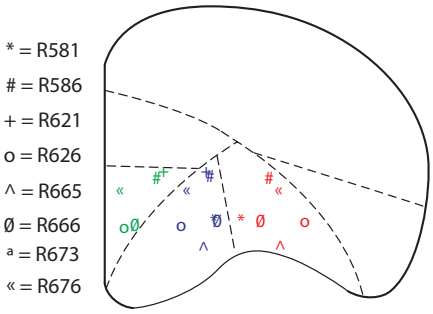

A
Time Deciding in Offer Zone
offer zone reaction time (s)
early late
1-5 s 1-10 s 1-30 s
B
Time Spent Lingering
linger zone reaction time (s)
early late
1-5 s 1-10 s 1-30 s
C
Offer and Wait Zone Decisions
decision probability
skips
quits
early late
1-5 s 1-10 s 1-30 s
D
Offer Zone (skips)
pSkip
offer delay (s)
E
Wait Zone (quits)
pQuit
offer delay (s)
F
LO: Cued Zone Entry Bursts
normalized fr
LZ
WZ
OZ
early late
1-5 s 1-10 s 1-30 s
MO
VO
LO
G
VO/MO: Offer Zone Ramping
activty slope across time in offer zone
early late
1-5 s 1-10 s 1-30 s
H
VO/MO: Offer Zone Delay Encoding
sensitivity to delay in oz
early late
1-5 s 1-10 s 1-30 s

Main Tables and Legends

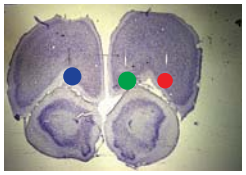
Table 1: Summary of which subregions within the OFC exhibited neural signatures during key task states on Restaurant Row. Larger circles indicate stronger effects.

		LO	VO	MO
Reward	Anticipation	●	•	
	Delivery	●	•	
	Consumption		●	●
Offer Zone Evaluations	Entry Burst	●		
	Short-Delay Activity (Accepts)	●		
	Long-Delay Ramping Activity (Skips)		●	•
Wait Zone Reevaluations	Entry Burst	●		
	Time Remaining	●	•	
	Time Spent	●		●
	Pre-Quit Activity		●	●

Figure S1. Histology. Probe tracks for LO, VO, and MO targets are depicted across rats, along with images from two example rats.



R586



R666

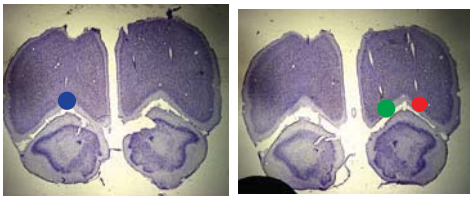


Figure S2. Distributions of reward-specific representations across the OFC during anticipation, delivery, and consumption. To quantify the degree to which ensemble activity encoded specific features of reward, we measured the mean difference in the probability of decoding to the current reward relative to each of the other three rewards. We compared the distribution of these scores to a shuffled distribution which randomly assigned reward flavors to current, next, past, and other rewards (see Methods for details). Positive values indicate information about specific rewards, and more positive values indicate a greater degree of information about current reward. (A-C) The distribution of current reward decoding scores in the LO (left panel), VO (middle panel) and MO (right panel) during reward anticipation (a), delivery (B), and consumption (C). Shuffled distributions are in black. Vertical lines reflect medians of the real and shuffled distributions. Stats performed on these distributions are reported in the main text.

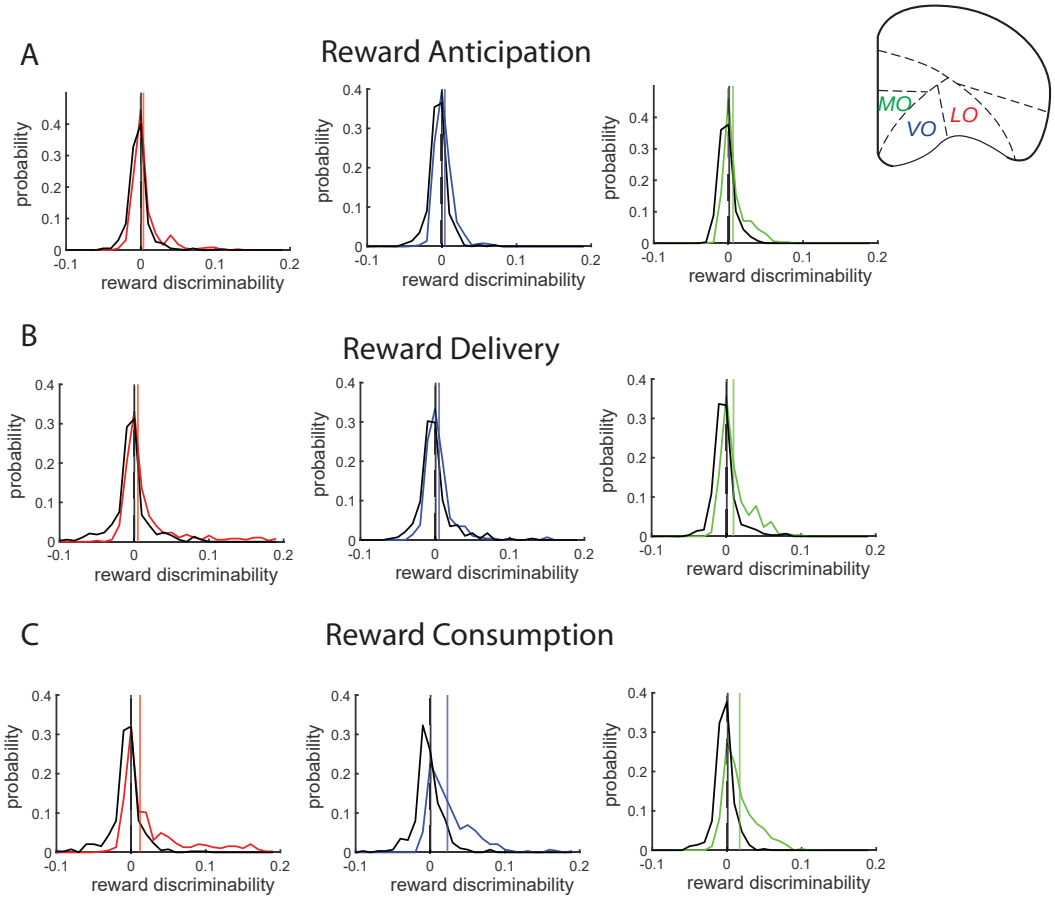

A
Reward Anticipation
B
Reward Delivery
C
Reward Consumption
probability
reward discriminability
MO
VO
LO

Figure S3. Decision-related activity in the offer zone across subregions. (A-C) Mean (± SEM) normalized firing rates during accept and skip encounters in LO (A), VO (B), and MO (C). The left panels show normalized activity oriented with respect to offer zone enter time while the right panels show normalized activity at offer zone exit time. (D) Cumulative density functions of the proportion of time in the offer zone in which LO activity exceeded VO activity for accept and skip encounters. Each panel shows these cumulative density functions broken up by the amount of time (reaction time – RT) spent in the offer zone.

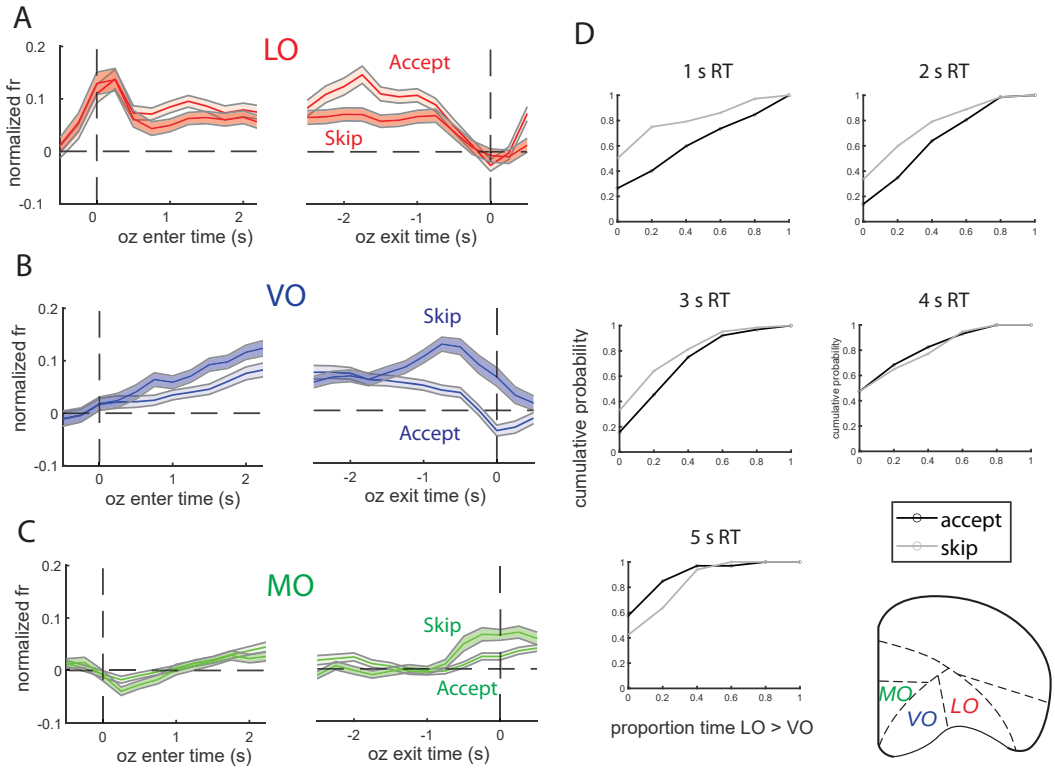

A
LO
Accept
Skip
normalized fr
oz enter time (s)
oz exit time (s)
B
VO
Skip
Accept
C
MO
Skip
Accept
D
1 s RT
2 s RT
3 s RT
4 s RT
5 s RT
cumulative probability
proportion time LO > VO
accept
skip
MO
VO
LO

Figure S4. Decision-related activity in the wait zone across subregions. (A-C) We measured the difference in neural activity between quit and stay encounters separately for medium (11-20 s) and long (21-30 s) delays. Mean (± SEM) difference in quit from stay normalized firing rate, oriented with respect to wait zone exit, in the LO (A), VO (B), and MO (C) averaged across medium and long delays. (D) To explore the possibility that heightened LO activity at wait zone entry during quits resulted from abnormally high LO activity for long-delay offers in the offer zone, resulting in an erroneous accept, we compared LO activity for quits, skips, and earns separately for medium and long delays. Mean (± SEM) LO activity during the epoch surrounding offer zone exit for encounters that were quit, skipped, or earned during medium and long delay offers. (E) Tuning curves of normalized LO firing rate to offer delay during the epoch surrounding offer zone exit when the offer was skipped, earned, and quit. (F) Quit probability as a function of time remaining and time spent across offer delays.

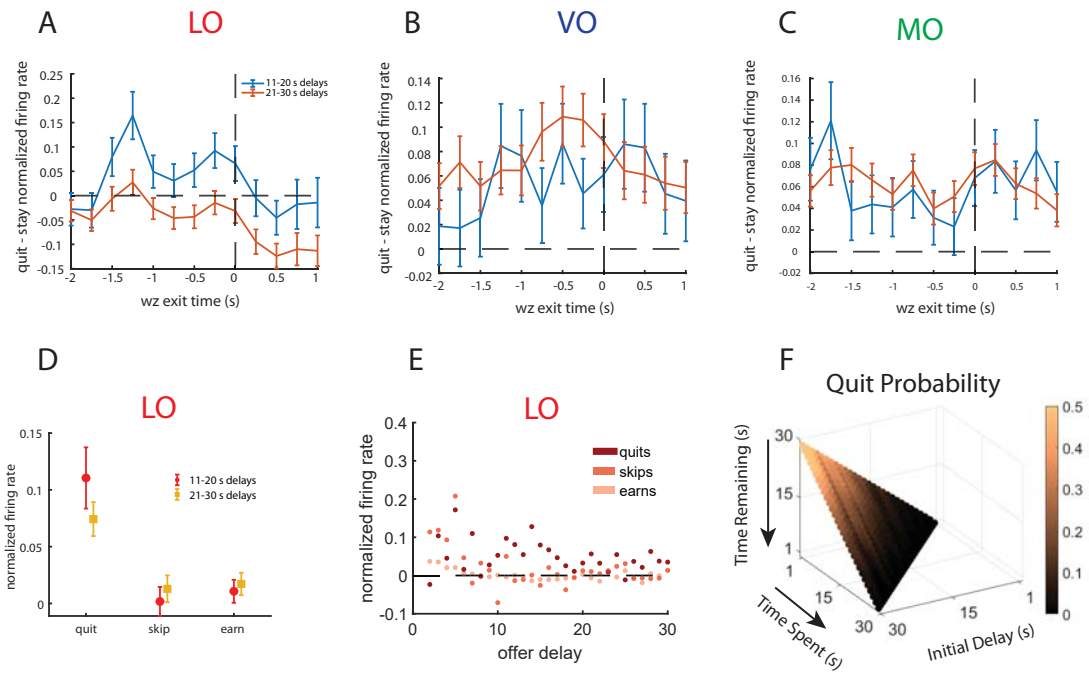

A
LO
B
VO
C
MO
quit - stay normalized firing rate
wz exit time (s)
11-20 s delays
21-30 s delays
D
LO
normalized firing rate
quit
skip
earn
E
LO
normalized firing rate
offer delay
quits
skips
earns
F
Quit Probability
Time Remaining (s)
Time Spent (s)
Initial Delay (s)

Figure S5. LO and VO neural dynamics were distinct between offer zone and wait zone decisions. (A-C) Mean (±SEM) normalized firing rate plotted at OZ entry (A), WZ entry (B), and reward delivery (C) for LO cells sensitive to OZ entry (blue), WZ entry (orange), and reward delivery (yellow). (D-F) Scatter plots of normalized activity scores for LO cells. (D) Mean normalized activity of LO cells at OZ (x-axis) and WZ (y-axis) entry, (E) at OZ entry (x-axis) and reward delivery (y-axis), (F) and at WZ entry (x-axis) and reward delivery (y-axis). Black = not sensitive to either zone entry event; blue = sensitive to OZ entry; orange = sensitive to WZ entry; yellow = sensitive to reward; purple = sensitive to events that define both axes. (G) Probability distribution of within-cell correlations of tuning curves to delay in the offer and wait zone within the LO. Correlations were normalized relative to a shuffled distribution in which the relation between normalized firing rate and delay were randomized. (H-N) Equivalent layout to (A-F), but for VO cells sensitive to skipping in the offer zone (blue), quitting in the wait zone (orange), or leaving in the linger zone (yellow).

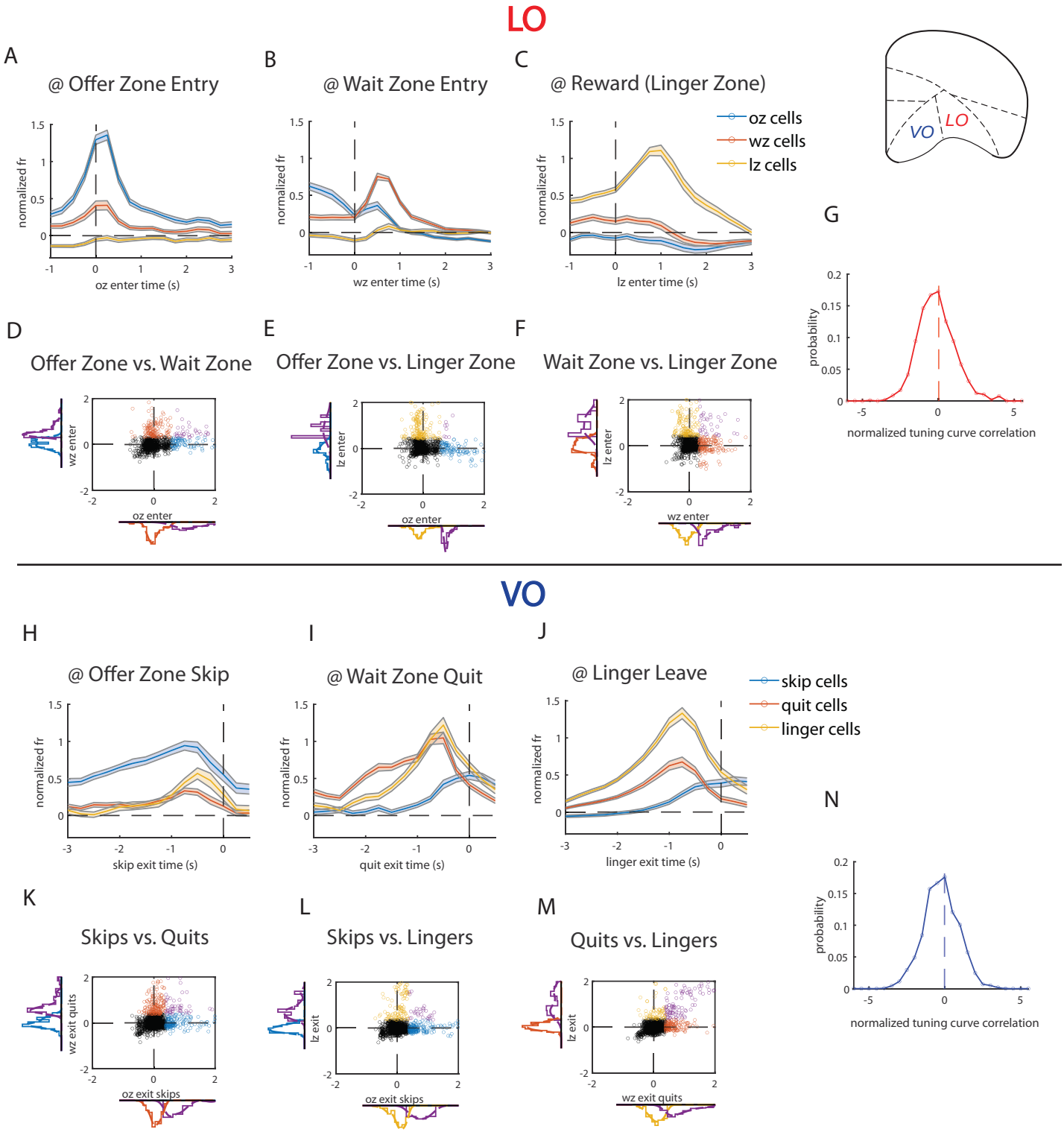

LO
A
@ Offer Zone Entry
normalized fr
oz enter time (s)
B
@ Wait Zone Entry
wz enter time (s)
C
@ Reward (Linger Zone)
oz cells
wz cells
lz cells
lz enter time (s)
VO
LO
G
probability
normalized tuning curve correlation
D
Offer Zone vs. Wait Zone
wz enter
oz enter
E
Offer Zone vs. Linger Zone
lz enter
oz enter
F
Wait Zone vs. Linger Zone
lz enter
wz enter
VO
H
@ Offer Zone Skip
skip exit time (s)
I
@ Wait Zone Quit
quit exit time (s)
J
@ Linger Leave
skip cells
quit cells
linger cells
linger exit time (s)
N
probability
normalized tuning curve correlation
K
Skips vs. Quits
wz exit quits
oz exit skips
L
Skips vs. Lingers
lz exit
oz exit skips
M
Quits vs. Lingers
lz exit
wz exit quits

Figure S6. Sequential decisions and OFC activity developed with experience in a reward-scarce environment. (B-B) Mean probability of quitting (A) and skipping (B) as a function of offer delay across recording sessions. Note that the first 5 sessions correspond to the 1-5 s condition whereas the next 5 sessions correspond to the 1-10 s condition. (C-D) Mean (± SEM) normalized activity at offer zone entry (left panels) and exit (right panels) for VO (C) and MO (D) ensembles during the four learning stages. (E-G) Mean (± SEM) normalized activity in LO ensembles at offer zone entry (E), wait zone entry (F), and reward delivery (G) across each of the four learning stages.

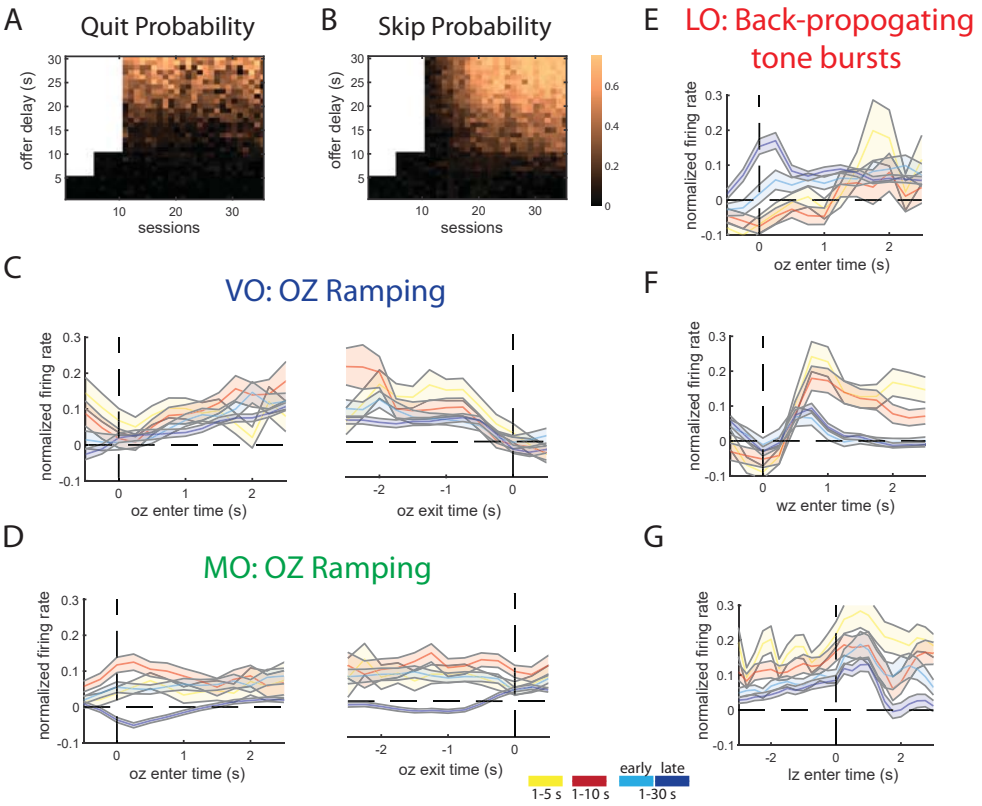

A
Quit Probability
offer delay (s)
sessions
B
Skip Probability
offer delay (s)
sessions
E
LO: Back-propogating tone bursts
normalized firing rate
oz enter time (s)
C
VO: OZ Ramping
normalized firing rate
oz enter time (s)
oz exit time (s)
F
normalized firing rate
wz enter time (s)
D
MO: OZ Ramping
normalized firing rate
oz enter time (s)
oz exit time (s)
G
normalized firing rate
lz enter time (s)
1-5 s
1-10 s
early
late
1-30 s

Table S1: Summary of recording information and cell counts across rats.

	Sex	Probe Type (# of channels)	Subregion	Hemisphere	Cell Count
R581	F	F-probe (64)	LO	Right	213
			VO	Both	567
			MO	Left	362
R586	M	E2-probe (32)	LO	Left	214
			VO	Right	593
			MO	Left	733
R621	F	E2-probe (32)	LO	Left	132
			VO	Right	164
			MO	Left	84
R626	F	E2-probe (32)	LO	Left	60
			VO	Right	115
			MO	Left	508
R665	M	E2-probe (32)	LO	Right	0
			VO	Left	0
			NA	NA	NA
R666	M	E2-probe (32)	LO	Left	31
			VO	Right	787
			MO	Left	782
R673	M	E2-probe (32)	LO	Right	667
			VO	Left	0
			MO	Right	0
R676	F	E2-probe (32)	LO	Right	111
			VO	Left	0
			MO	Right	0