Behavioral and neurophysiological correlates of regret in rat decision-making on a neuroeconomic task

Adam P. Steiner[1] and **A. David Redish**[2]

[1]Graduate Program in Neuroscience, University of Minnesota, Minneapolis, MN, USA

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

Summary

Disappointment entails the recognition that one did not get the value one expected. In contrast, regret entails the recognition that an alternate (counterfactual) action would have produced a more valued outcome. Thus, the key to identifying regret is the representation of that counterfactual option in situations in which a mistake has been made. In humans, the orbitofrontal cortex is active during expressions of regret, and humans with damage to the orbitofrontal cortex do not express regret. In rats and non-human primates, both the orbitofrontal cortex and the ventral striatum have been implicated in decision-making, particularly in representations of expectations of reward. In order to examine representations of regretful situations, we recorded neural ensembles from orbitofrontal cortex and ventral striatum in rats encountering a spatial sequence of wait/skip choices for delayed delivery of different food flavors. We were able to measure preferences using an economic framework. Rats occasionally skipped low-cost choices and then encountered a high-cost choice. This sequence economically defines a potential regret-inducing instance. In these situations, rats looked backwards towards the lost option, the cells within the orbitofrontal cortex and ventral striatum represented that missed action, rats were more likely to wait for the long delay, and rats rushed through eating the food after that delay. That these situations drove rats to modify their behavior suggests that regret-like processes modify decision-making in non-human mammals.

Introduction

Regret is a universal human paradigm[1–5]. The experience of regret modifies future actions[1,4,6]. However, the experience of regret in other mammals has never been identified; it is not known whether non-human mammals are capable of experiencing regret. Although non-human animals cannot verbally express regret, one can create regret-inducing situations and ask whether those regret-inducing situations influence neurophysiological representations or behavior: Do non-human animals demonstrate the neural correlates of regret in potential regret-inducing situations?

When evaluating the experience of regret, it is important to differentiate regret from disappointment[4,7,8]. Disappointment is the realization that a realized outcome is worse than expected[7,8]; regret is the realization that the worse than expected outcome is due to one's own mistaken action[1–3,9]. Disappointment can be differentiated from regret through differences in the recognition of alternatives[2,6,8,10]. Regret can be defined as the recognition that the option taken resulted in a worse outcome than an alternative option/action would have. The revaluation of the previous choice in context of the current choice is the economic foundation of regret [4,6].

Humans with damage to the orbitofrontal cortex (OFC) do not express regret[2], and fMRI experiments reveal activity in the orbitofrontal cortex during regret[1,11]. In rats and non-human primates, the OFC has been implicated in decision-making, particularly in the role of expectations of future reward and the complex calculations of inferred reward[12–14,15–17]. Orbitofrontal cortical neurons represent the chosen value of an expected future reward[14,18,19], and an intact OFC is critical for reversal learning[20,21](recent evidence suggests that OFC may have a more specialized role and is not necessary for reversal learning at least in primates[22]). Orbitofrontal cortex has been hypothesized to be critical for learning and decision-making[10,15,23,24], particularly in the evaluation of expected outcomes[14,25].

The ventral striatum (vStr) has also been implicated in evaluation of outcomes[26–29], particularly in evaluation during the process of decision making[23,29,30]. Neural recordings vStr and OFC in rats have found representations of reward, value and prediction of expected value in both structures[12,25,29,31–33]. In the rat, lesion studies suggest orbitofrontal cortex is necessary for recognition of reward-related changes that require inference, such as flavor and kind, while vStr is necessary for recognition of any changes that affect value[15,23]. In rats deliberating at choice points, vStr reward representations are transiently active before and during the reorientation process[29], but reward representations in OFC are only active after the reorientation process is complete[25].

We developed a neuroeconomic spatial decision-making task for rats (Restaurant Row) in which the rat encounters a serial sequence of take/skip choices. The Restaurant Row task consisted of a large inner loop, approximately one meter in diameter with four spokes proceeding out from the inner loop (Fig 1a). Each zone supplied a different flavor of food (banana, cherry, chocolate, and unflavored). Flavor locations remained constant throughout the experiment. Rats were trained to run around the loop, making stay/skip decisions as they passed each spoke.

Upon entering each zone, rats encountered different offers of delays. Zone entries were defined entirely by the detected position of the rat's head and were not explicitly marked on the track. On entry into a zone, a tone sounded, the pitch of the tone indicated the delay the rat had to wait in order to receive reward (higher pitch = longer delay). As long as the rat stayed within the zone, the delay counted down, with each subsequent second indicated by a lower pitch tone. If the rat left the zone, the countdown stopped, no sound was played, and the offer rescinded – the rat's only option was to proceed on to the next spoke and the next zone.

The delays were independently selected pseudo-randomly from a uniform distribution ranging from 1–30s (for two rats) or 1–45s (for two rats). The delay offered at each zone encounter was independent of other zones for that lap. When making a decision to stay or skip at a given zone (when offered a given delay), the only information the rat had was the flavor of the food offered (flavor locations remained constant throughout the experiment), the delay it would have to wait (delay signaled by pitch of the auditory cue), and the probability distribution of any future offers (offers were drawn from a uniform distribution of 1–30s or 1–45s).

Rats were run for one 60 minute session per day. This time-limit meant that rats had a "time-budget" of 60 minutes to spend foraging for food. Because the session was time-limited, the decision to stay or skip a zone was not independent of the other zones. Waiting at one zone was time that could have been spent at another zone. An economically-maximizing rat should distribute its time between the offers, waiting for valuable offers, but skipping expensive offers. Assuming that an animal likes some flavors more than others, the economic value of an offer should depend on the delay offered and the animal's preferences.

Results

Revealed preferences

Four rats were trained on the Restaurant Row task (Fig 1). Thresholds and preferences were determined by utilizing an economic framework. All four rats showed similar behaviors in that they were likely to wait through the delay for delays less than a threshold, but unlikely to wait through the delay for delays greater than a threshold. Skips and stays were evenly distributed throughout each session. When rats skipped an option, they left within the first ~5s, independent of delay (Supplemental Fig S1). The threshold between waiting and skipping tended to be different for the different flavors for a given rat (Fig 1b, Supplemental Fig S1). These thresholds were consistent within rat, but differed between rats (Fig 1c–e), indicating an underlying revealed, economic preference for each flavor of food that did not change across a session (Supplemental Fig S2). There were no differences in reward handling between delays, rats generally waited 20–25s after consuming reward before leaving for the next zone (Supplemental Fig S3).

In order to directly test whether the rats were making economic decisions (comparing value and cost), after completing the primary Restaurant Row experiment, we ran two of the rats on a variant of the task in which one reward site provided three times as much food as the other three sites. In this control task, rats were run on four 20-minute blocks, so that each site could be the large reward site for one block. (The order of which reward site provided excess reward was varied pseudo-randomly. Rats were removed to a nearby resting location for one minute between blocks.) Rats were consistently willing to wait longer for more food (Supplemental Fig S4). All results reported here except for Supplemental Fig S4 are from the primary Restaurant Row experiment.

Reward responses

We recorded 951 neurons from orbitofrontal cortex (OFC) and 633 neurons from ventral striatum (vStr) (see Supplemental Fig S5 for recording locations). Neurons were identified as reward-responsive if their activity during the 3s following reward delivery was significantly different (p<0.05, Wilcoxon) than a bootstrapped (n=500) sample of activity during 3s windows taken randomly across the entire session[25,29]. 81% of OFC neurons responded to reward; 86% of vStr neurons responded to reward. Responses in both OFC and vStr often differentiated between the four reward sites. (Supplemental Figures S6, S7).

Because responses differentiated between rewards, a decoding algorithm applied to these neural ensembles should be able to distinguish between the reward sites. We used a Bayesian decoding algorithm with a training set defined by the neuronal firing rate in the 3s following delivery of reward (which we used to calculate *p(spikes|Reward)*) or a training set defined by the neuronal firing rate in the 3s following entry into a zone (which we used to calculate *p(spikes|Zone)*). In order to provide a control for unrelated activity, we also included a fifth condition in our calculation, defined as average neuronal firing rate during times the animal was not in any countdown zone. Thus, the training set consisted of five expected firing rates: firing rate after reward-receipt or zone entry (1) at banana, (2) at cherry, (3) at chocolate, and (4) at nonflavored, plus a fifth control of expected firing rate (5) on the rest of the maze. From this training set, Bayesian decoding uses the population firing rate at a given time to derive the posterior probability of the representation *p(Reward|spikes)* or *p(Zone|spikes)*. For simplicity, we will refer to these two measures as *p(Reward)* and *p(Zone)*.

In order to pool data from all four sites, we categorized and rotated each reward site based on the current position of the animal. This gave us four sites that progressed in a serial manner – the previous site, the current site, the next site, and the opposite site (Fig 2c). All analyses were based on this categorization. All analyses used a leave-one-out approach so that the encounter being decoded was not included in the definition of the training set.

Both OFC (Fig 2a) and vStr (Fig 2b) were capable of reliably distinguishing between the current reward site (Fig 2c) and the other sites (Supplemental Fig S8). Shuffling the interspike intervals of the cells removed all of these effects. *p(Reward)* and *p(Zone)* calculated from shuffled data were consistently at 0.14 (Supplemental Fig S8).

Zone entry responses

Previous research has suggested that in simple association tasks in which cues predict reward, both OFC and vStr cells respond to cues predictive of reward[12,15,24,28,31,34,35]. Both OFC and vStr neural ensembles distinguished the different zones both at the time of reward (Fig 2a–c) and at the time of entering the zone/cue onset (Fig 2d–f) (single cell differentiation see Supplemental Fig S6 and S7, decoding differentiation Supplemental Fig S8 and S9). These representations were related; thus, neural activity in OFC and vStr also predicted the reward type of the current zone during zone enter/cue onset (Fig 2g–i, Supplemental Fig S10). Shuffling the interspike intervals eliminated these effects (Supplemental Fig S11).

Both OFC and vStr responded strongly under conditions in which the animal determined the cost to be worth staying (e.g. when the delay was below threshold, Fig 3a,b). In contrast, neither structure represented expectations of reward under conditions in which the animal determined the cost to not be worth it (i.e. skips, when delay was above threshold, Fig 3c,d). This suggests that these structures were indicating expected value, and predicting future actions. To directly test this hypothesis, we compared reward-related decoding when the rat encountered a delay near threshold (threshold +/−2 s) and either stayed to sample the feeder (Supplemental Fig S12a,b) or skipped to proceed to the next reward option (Supplemental Fig S12c,d). When the animal stayed (waiting for a reward), both OFC and vStr increased their representations of the current reward at the time of zone entry. In contrast, when the animal skipped the current reward, neither OFC nor vStr reliably represented the current reward/zone. Shuffling the interspike intervals of the cells removed these effects (Supplemental Fig S13).

Regret

Regret entails the recognition that one has made a mistake, that an alternate action would have been a better option to take[4,6]. As noted above, a regret-inducing situation requires two properties be satisfied: (1) the undesirable outcome should be a result of the agent's previous action, and (2) following the selection of an option, the outcome/value of all options needs to be known, including the outcome/value of the unselected options. Our task and behavior satisfies these conditions. Because the rats were time-limited on the Restaurant Row task, encountering a high-cost delay after not waiting through a low-cost delay means that skipping the low-cost delay was a particularly expensive missed opportunity.

In the Restaurant Row task, a rat would sometimes skip an offer that was less than that rat's threshold for that flavor on that day and then encounter an offer at the subsequent site that was greater than that rat's threshold for that flavor on that day. Because the delay is a cost and value is matched (by definition) at threshold, this sequence is one in which the rat skipped a low-cost offer, only to find itself faced with a high-cost offer. From the economic and psychology literature, we can identify these sequences as potential "regret inducing" situations[4,6]. We can compare these conditions to control conditions which we would expect to induce disappointment rather than regret.

Literature suggests that during regret, there should be manifest changes in the animal's behavior and neurophysiology that reflect recognition of the missed opportunity, as well as subsequent behavioral choices that one might not have made normally. Theoretically, the key to regret is a representation of the action not taken[3,5,9,36,37]. This implies that there should be representations of the previous choice during the regret-inducing situations, particularly in contrast to control conditions that are "merely disappointing".

Thus, we define a <u>regret-inducing situation</u> as one in which (1) the rat skipped a low-cost/ high-value reward (delay less than measured threshold for that flavor for that day), and then (2) the rat encountered a high-cost/low-value reward (delay greater than measured threshold for that flavor for that day). In this situation, the rat has made an economic mistake: if it had taken a different action (waited for that previous reward), it would have had a more valuable

session. For consistency, we will refer to the opportunity in (1) as the <u>previous zone/ previous reward</u> and the opportunity in (2) as the <u>current zone/current reward</u>.

As noted above, one needs to differentiate regret-inducing sequences from sequences that are merely disappointing. By definition, a disappointing sequence occurs when one encounters a situation that is worse than expected, but not due to one's own agency. There are two controls that need to be taken into account, a control for the sequence of offers (control 1), and a control for the animal's actions (control 2).

To control for the sequence of offers, we took sequences in which the rat encountered the same sequence of offers, but took (stayed for) the first offer. This matched control should only induce disappointment (worse than expected, but not due to the fault of the rat)[7,8]. Control 1 differs from the regret-inducing situation only in that the rat took the previous offer rather than skipping it. In summary, control 1 was defined as situations in which the delay at the previous zone was below threshold and the rat waited for reward, followed by an encounter at the current zone such that delay was above threshold. In this situation, the rat did not make a mistake (as it waited for reward at the previous zone); the delay at the current zone was merely worse than the rat was willing to wait for, making the rat (presumably) disappointed. Control 1 controls for the sequence observed by the rat.

To control for the rat's actions, we took sequences in which the offer at the previous zone was greater than threshold (and skipped), and, again, the rat encountered a higher-than-threshold offer at the current zone. In this second control condition, the rat skipped the previous offer, but that was the "correct" action to take, as the previous offer was above threshold. This second control condition should also induce disappointment because the rat has encountered two high-cost offers in a row. But this second control condition should not induce regret, because the rat's actions were consistent with its revealed preferences. Control 2 differs from the regret-inducing situation only in that the delay at the previous offer was above rather than below threshold. In summary, control 2 was defined as situations in which the delay at the previous zone was above threshold, followed by an encounter at the current zone such that delay was above threshold. In this situation, the rat did not make a mistake (since it skipped a high-cost delay at the previous zone), but the delay at the current zone was worse than the rat was willing to wait for, making the rat (presumably) disappointed. Control 2 controls for the reward sequence seen by the rat.

Potential regret and control instances were found within each session by comparing the delays at each of the zones to the threshold of that zone for that rat for that day. Regret instances and control instances were evenly distributed throughout each session across all rats. The distribution of the high-cost offers at the current zone did not differ between the potential regret-inducing sequences and matched controls (Supplemental Fig S14).

Behaviorally, rats paused and looked backwards towards the previous option upon encountering a potentially regret-inducing sequence, but did not do so in either control condition (Fig 4). We identified pause-and-look events as points of high curvature and derived an orientation (see **Methods**). During potential regret-inducing sequences, rats were more likely to look backwards towards the previous option (Fig 4d) than during either of the

matched control conditions (Fig 4e,f) (p<0.05, Watsons Circular U Test). In the first control condition (where the rat took a good offer and then encountered a bad offer), the rat tended to look towards the current zone, but then skip it and go on to the next zone. In the second control condition (where the rat encountered two bad offers in a row), the rat tended to look towards the next zone. Thus, there was a behavioral difference, implying that the rats recognized these three situations differently.

During potential regret instances individual reward-responsive neurons in OFC and vStr showed activity patterns more consistent with the previous reward than the current one (Fig 5). Neural activity peaked immediately following the start of the look-back towards the previously skipped, low-cost reward. In order to quantify these changes in representation during regret-inducing situations and disappointment-inducing controls, we examined the population dynamics using a Bayesian decoding algorithm. Population decoding analyses offer insight into the dynamics of neural population. Ensemble activity more accurately represents the dynamics of the entire population compared to that of a single cell. In order to determine the neural population representation during these situations, we measured the Bayesian representations of *p(Reward)* and *p(Zone)* from the ensemble including all cells.

While our first inclination was to look for representations of the missed reward, human subjects self-report that they regret actions taken or not, more than they do missed outcomes[3,5,9,36]. We did find a weak representation of the missed reward (not significant, see Supplemental Fig S15). However, we found that there were strong representations of the previous decision-point (*p(Zone)*) that were significantly different than all other zones (Fig 6a–c, outside the 95% confidence interval as determined by empirical cumulative distribution function). This differentiation of the previous zone was not observed in either control condition. In the first control (same sequence), both OFC and vStr demonstrated increased representations of the next zone (Fig 6d–f). By definition, these control instances were high-cost encounters with the current reward site (e.g. above threshold), and, thus the rats were likely to skip them. In the second control condition (two bad offers), the representation of both the current and previous zones increased and were different than the representations of other rewards (Fig 6g–i). However, this response was markedly different from that seen during potential regret instances as the increase in representation of the previous zone could not be differentiated from the increase in representation of the current zone. Shuffling interspike intervals eliminated all of these effects (green traces Fig 6). Other more positive situations (rejecting a low-cost previous offer and then encountering a low-cost offer, or rejecting a high-cost offer and then encountering a low-cost offer) both led to strong representations of the current zone (Supplemental Fig S16). In addition, when rats stayed for an above threshold delay, but then encountered a below threshold delay (which could be described as a potential regret inducing condition) we again found increased representations of the previous zone (Supplemental Fig S17). The representations of the previous zone in this condition (Stay at Delay A > Threshold A to Delay B < Threshold B) were smaller when compared to the previously-described regret-inducing condition.

Thus, the rats showed different behaviors and different neurophysiological representations during regret-inducing situations, both of which reflected the information processing we would expect to see during regret. As noted above, an important role of regret in decision-

making is that it changes subsequent decisions[37–39]. Consistent with this hypothesis, we found that rats were more likely to stay at the high cost option in a regret-inducing situation than under either control condition (vs. first control condition, p=0.01; vs. second control condition p=0.06, Wilcoxon, Fig 7a). In addition, rats spent less time after eating the food before proceeding on to the next reward site following regret-inducing situations compared to non-regret conditions (Fig 7b). The handling time distributions were significantly different (regret condition versus all nonregret conditions, Wilcoxon, $p \ll 0.001$). After waiting for food through a long-delay in a regret-inducing situation, rats rushed through eating and quickly went on to their next encounter.

The hypothesis that the neural representation of the previous zone reflects some information processing related to regret implies that there should be a relationship between that representation of the previous zone and the animal's subsequent actions. The hypothesis predicts that a stronger representation of the previous zone would lead to an increased likelihood of taking the high-cost (current) offer. To determine if there was a relationship between a rat's willingness to take the high-cost offer and the neurophysiological representations, we compared the ratio of representations of the previous and the current zones and categorized these representations by stay/skip decisions at the current zone. As shown in Fig 8, this ratio was increased when the animal decided to stay, but only within the regret-inducing situations. The ratio was unrelated to the decision to stay in the two control conditions. In regret-inducing situations, animals were more willing to stay on trials in which they showed an increased representation of the previous zone relative to the current zone.

Discussion

Regret is the introspective recognition that a previously chosen action led to a less desirable outcome than an alternative action would have. The two keys to identifying regret are value and agency. The Restaurant Row task, in which rats made economic (value-related, cost-dependent) decisions allowed us to identify potentially regret-inducing situations. First, the Restaurant Row task was an economic task, in which rats revealed economic preferences just as human and non-human primates do[14,40,41]. Second, because the rats had a limited time-budget, encountering a bad (above-threshold) offer after skipping a good (below-threshold) offer, meant that the rat had missed an opportunity. By standard economic and psychological definitions, this sequence should induce regret[4,6,9]. We were able to identify two matched sets of controls that should induce disappointment but not regret: (1) situations in which the rat encountered a similar sequence of offers but took the previous low-cost option, and (2) situations in which the rat encountered two above-threshold offers and skipped the previous high-cost option.

Our data indicate that behavioral and neurophysiological differences between the potential regret-inducing situations and the controls were consistent with a hypothesis that the rats were expressing something akin to human regret. During the regret-inducing situation, rats looked backwards towards the previous (missed) goal, and the OFC and vStr were more likely to represent that previous goal. After it, rats were more likely to wait out the (current) high-cost offer, and rushed through handling their reward when they did. Interestingly, we

found that the neurophysiological representations of counterfactual information in the regret-inducing situation were more strongly related to the missed action (activity when the action was taken, measured by *p(Zone)*) than to the missed outcome (activity when the reward was received, measured by *p(Reward)*). This is consistent with data that humans express more regret about the actions taken (or not taken) than about the missed outcomes[3,5,9,36,37].

The Restaurant Row task had three features that made it particularly well suited to the identification of regret. First, it is an economic task on which rats reveal preferences. Second, the inclusion of four "restaurants" allowed us to differentiate a general representation of other rewards from a specific representation of the mistaken choice. We found a clear and significant representation of the previous (lost) zone, not the next or opposite zones. Third, the Restaurant Row task separates the choice of waiting (staying) or going (skipping) from reward-receipt. This separation allowed us to differentiate the regret-induced representation of the previous (lost) reward (a small effect) from the regret-induced representation of the previous (mistaken) action (a large effect). Regret is more about the things you did or did not do than about the rewards you lost [5,9,36].

Prior evidence indicates that rats can combine information to form an expectation of a novel reward (imagining a particular outcome), and a role for both OFC[16,17,42] and vStr (if a model in the evaluation steps of the task exists) in this process[23,24]. Our data indicates that violation of an expectation initiates a retrospective comparison (regretting a missed opportunity). As with the prospective calculation of expectation, this retrospective calculation of expectation influences future behavior – rats are more willing to wait for reward following a regret instance. These two processes, the act of imagining future outcomes and the process of regretting previous, poor choices, are both necessary to modify future decisions to maximize reward. While some evidence suggests that OFC represents economic value[14], the representation of regret is more consistent with the hypothesis that OFC encodes the outcome parameters of the current, expected or imagined state[15–17,23]. The data presented here is also consistent with the essential role OFC plays in proper credit assignment[43–45]. Previous studies have identified potential representations of the counterfactual could-have-been-chosen option in rats[25], monkeys[19], and humans[11]. In humans, representations of the value of the alternative outcome activate OFC[1,11]. Abe and Lee[19] found that there were representations of an untaken alternate option in monkey OFC on a cued-decision-making task, in which the alternate option that should have been taken was cued to the monkey after the incorrect decision.

The connectivity between OFC and vStr remains highly controversial with some evidence pointing to connectivity[46–48] and other analyses suggesting a lack of connectivity[49,50]. The anatomical and functional mechanisms through which the OFC and vStr derive their representations of regret-related counterfactual information remains unknown. In addition the analyses used here lack the temporal resolution necessary to determine any interactions between structures.

The Restaurant Row task introduced here allowed economic measures to identify potential regret-inducing situations, in which the rat made a decision that placed it in a less valuable

situation. Because the task was time limited, any decision to wait for a reward decreased the amount of time available to receive future rewards. Human subjects self-report that they regret actions taken or not, more than they do missed outcomes[3,5,9,36]. Intriguingly, our decoding results showed strong representations during regret-inducing situations of the previous zone-entry where the decision was made and the action taken *p(Zone)*, but weak and non-significant representations of the missed outcome *p(Reward)*. Most hypotheses suggest that the role of regret is a revaluation of a past opportunity that drives future behavioral changes[4,6].

After making a mistake and recognizing that mistake, rats were more likely to take a high-cost option and rush through the consumption of that less-valuable option.

Methods

Animals

Four Fisher Brown Norway rats aged 10–12 months at the start of behavior were used in this experiment. Rats were maintained at above 80% of their free-feeding weight. All experiments followed approved NIH guidelines and were approved by the Institutional Animal Care and Use Committee at the University of Minnesota. Four rats is a standard sample size for behavioral neurophysiology experiments measuring information processing in large neural ensembles. Each rat was from a different litter.

Experimental Design

The Restaurant Row task consisted of a central ring (approximately three feet in diameter) and four spokes leading off of that ring (Fig 1a). At the end of each spoke, a feeder (MedAssociates, St. Albans VT) dispensed two 45mg food pellets of a given flavor (*banana, cherry, chocolate, and unflavored[plain],* Research Diets, New Brunswick, NJ). A given flavor remained at a constant spoke through the entire experiment. As the rats proceeded around the track, the rat's position was tracked from LEDs on the head via a camera in the ceiling. A spatial zone was defined for each spoke that included the complete spoke and a portion of the inner loop and aligned with the inner loop such that a rat could not miss a zone by running past it (boxes in Fig 1a). The zone's entries were separated by ninety degrees and each one led to a potential reward location approximately half a meter from the entry point on the central, circular track. A trigger <u>zone</u> was defined so as to include a spoke and the portion of the ring nearby. Zones were primed in a sequential manner so that the rat ran in one direction around the loop. When the rat entered a primed zone, a tone sounded indicating the delay the rat would have to wait in that zone to receive food. Offered delays ranged from 1 second (identified by a 750Hz tone) to 45 seconds (12 kHz). As long as the rat remained in the active zone, a tone sounded each second, decreasing in pitch (counting down by 250Hz increments). If the rat left the zone, the tones stopped, and the next zone in the sequence was primed. In practice, rats waiting out a delay would proceed down the spoke and wait near the feeder; rats skipping a zone would proceed directly on to the next trigger zone. Each rat ran one 60 minute session each day. Zones were defined as a box around each reward location and extended onto the circular inner portion of the maze such that a rat was required to pass through a zone. Each reward arm extended approximately 22

inches away from the circular portion of the track. During training, rats were allowed to run the task in any manner they saw fit. However, rewards were only available if they traveled through the zones sequentially, Zone1 to Zone2, to Zone3, to Zone4. If a rat were to travel backwards the rat would have had to complete approximately three laps in order to prime the previous zone. Rats quickly learned that this behavior was not viable. Within seven days, rats learned to travel in only one direction and to pass through each zone sequentially.

Rats were initially handled and accustomed to the different flavors as described previously[25]. Rats were shaped to the task in three stages. In the first stage, all offers were 1 second. Once rats ran 30 laps/session consistently, they progressed to the second stage. In the second stage, each offer was randomly chosen from 1 to 10 seconds (uniform distribution, independent between encounters). Again, once rats ran 30 laps/session consistently, they progressed to the third stage, in which they faced the full Restaurant Row task with offers selected between 1 and 30 seconds (uniform distribution). Two rats often waited out the full 30 seconds at some locations, so delays were increased for those rats to range from 1 to 45 seconds.

Once rats were completing at least 50 laps/session on the full Restaurant Row task, they were implanted with hyperdrives targeting the ventral striatum and orbitofrontal cortex. Rats were then re-introduced to the task until running well. Each day, rats were allowed to run for 60 minutes and often completed upwards of 70 entries per zone. Rats received all of their food on the track each day.

Control task (4×20)—In order to confirm the economic nature of the Restaurant Row paradigm, two rats ran an additional task after completing all recordings. In this modified version, each rat ran one session per day which consisted of four blocks of 20 minutes/block. In each block, one reward site provided three food pellets (of its corresponding flavor), while the other three reward sites provided one food pellet (of their corresponding flavors). Delays ranged from 1 to 45 seconds (uniform distribution). Each of the four sites were the "3-pellet" site for one of the four blocks each day. Which site was improved in which block was pseudo-randomly varied across days. Rats were removed to rest on a nearby flower pot for 60 seconds between each block.

Surgery—Rats were implanted with a dual bundle 12 tetrode + 2 reference hyperdrive[25,29] aimed at ventral striatum (6 tetrodes + 1 reference, M/L +1.8mm, A/P +1.9mm) and orbitofrontal cortex (6 tetrodes + 1 reference, M/L +2.5mm, A/P +3.5mm). For two rats, the two targets were left vStr and left OFC, while for two rats the two targets were right vStr and right OFC. Following surgery, tetrodes were turned daily until they reached vStr and OFC. Upon acquisition of large neural ensembles and a return to stable behavior on the maze, each rat ran a minimum of 10 recording days. Data reported here came from a total of 47 sessions distributed evenly over the four rats, R210: 12 sessions, R222: 12 sessions, R231: 13 sessions, R234: 10 sessions (Supplemental Table 1)

Data Analysis

No data that met the inclusion critera (as defined in the main text) were excluded. Analyses were automated and applied uniformly to all instances meeting the inclusion criteria.

Analyses were based on an encounter by encounter basis. Clusters were cut on a session by session basis; experimenters were blind to behavior when cutting clusters.

Behavior

Threshold calculation—At every encounter with a reward zone, the rat could <u>wait</u> through the delay or <u>skip</u> it and proceed to the next zone. If the rat chose to skip, it tended to do so quickly (Supplemental Fig S1). As can be seen in **Main Text** Fig 1, rats tended to wait for short delays and skip long delays, as expected. In order to determine the threshold, we defined stays as 1 and skips as 0 and fit Sigmoid functions of stay/skip as a function of delay using a least squares fit (Matlab, MathWorks, Natick MA). The threshold for "above/below" calculations was defined as the mid-point of the Sigmoid. We determined a threshold for each rat for each session for each zone. All preference data was measured during the task and each rat demonstrated a different preference indicated by the amount of time that rat was willing to wait for reward.

Identifying regret-inducing and control situations—On entry into a given ("current") zone, we defined the situation as <u>regret-inducing</u> if it met the following three conditions: (1) the offer at the previous zone was a delay < threshold for that previous zone for that rat for that session. (2) The rat skipped the previous offer. (3) The offer at the current zone was a delay > threshold for that current zone for that rat for that session.

The first control was defined using the same criteria as for regret-inducing situations, but that (2) the rat took the previous offer. This control situation keeps the sequence of offers the same, but controls for the rat's agency/choice.

The second control was defined using the same criteria as for regret-inducing situations, but that (1) the offer at the previous zone was a delay > threshold for that previous zone for that rat for that session. This control situation keeps the rat's choices the same, but makes the choice to skip the previous option the correct one (see Supplemental Table 2 for a summary of conditions).

Curvature—In order to identify the pause-and-look behavior, we measured the curvature of the path of the animal's head, and identified the point of maximum curvature and the direction of that point. Curvature was measured through the following algorithmic sequence: the position of the head was measured at 60 Hz from the LEDs on the headstage via the camera in the ceiling, giving $<x,y>$ coordinates, velocity $<dx,dy>$ was calculated using the Janabi-Sharifi algorithm[51] acceleration $<ddx, ddy>$ was calculated by applying the Janabi-Sharifi algorithm to $<dx,dy>$. Finally, the curvature[52] at each moment was defined as

$$\frac{dx*ddy+dy*ddx}{(dx^2+dy^2)^{1.5}}$$

Neurophysiology

Cells were recorded on a 64 Channel Analog Cheetah-160 Recording system (Neuralynx, Bozeman MT) and sorted offline in MClust 3.5 (Redish, current software available at http://

redishlab.neuroscience.umn.edu/). For all sessions, the position of the rat was tracked via overhead camera viewing colored LEDs on the headstage.

Reward Responsiveness—We are interested in determining how a cell modulates its activity during reward delivery. To measure this quantitatively, we compared the firing rate of the cell in the 3s after reward-delivery to 500 randomly-selected 3s intervals throughout the task. If a cell's firing rate is different (whether increased or decreased) during reward delivery, then it carries information about reward delivery. We can measure this change by determining if the activity during the 3s after reward-delivery was significantly different than the bootstrap. Because these distributions were not Normal, we used a Wilcoxon to calculate significance. Responsiveness to each reward site was calculated independently (see Supplemental Table 1 for summary of cells per rat).

Bayesian decoding—We used a Bayesian decoding algorithm[53] with a training set defined by the neuronal firing rate at specific times of interest (250 ms window). Any decoding algorithm consists of three parts: (1) a training set of tuning curves which defines the expected activity as a function of the variable in question, (2) a test set of spikes or firing rates, and (3) the posterior probability calculated from (1) and (2). In this manuscript, we used two decoding processes – one in which the tuning curves were defined as the neural activity in the 3 seconds after reward-delivery at the four reward-locations [*p(Reward)*], and one in which the tuning curves were defined as the neural activity in the 3 seconds after initial cue-delivery (zone entry) [*p(Zone)*]. When calculating *p(Zone)*, time after reward delivery was not included. This was only important for delays < 3 seconds.

p(Reward)

Throughout the paper we refer to this measure as "*p(Reward)*", however, mathematically, it is *p(reward/spikes)*. Assuming a uniform distribution of reward priors, this equation is:

$$p(reward|spikes) = \frac{p(spikes|reward) * 0.20}{p(spikes)}$$

We defined the training set of *p(spikes/reward)* as the firing rate during the 3s after a given reward delivery (e.g. *p(spikes/banana), etc.*). In order to provide a control for unrelated activity, we also included a fifth condition in our calculation, defined as average firing rate during times the animal was not in any countdown zone. Thus, the training set consisted of five expected firing rates: firing rate after reward-receipt (1) at banana, (2) at cherry, (3) at chocolate, and (4) at nonflavored, plus a fifth control of expected firing rate (5) on the rest of the maze. Because of the inclusion of the fifth (average firing rate when not at reward) condition, the normalization factor is 0.20.

p(Zone)

Throughout the paper we refer to this measure as "*p(Zone)*", however, mathematically, it is *p(zone/spikes)*. Assuming a uniform distribution of reward priors, this equation is:

$$p(zone|spikes) = \frac{p(spikes|zone)*0.20}{p(spikes)}$$

We defined the training set of *p(spikes/zone)* as the firing rate during the 3s after entry into a given trigger zone (e.g. *p(spikes/banana-zone), etc.*). In order to provide a control for unrelated activity, we also included a fifth condition in our calculation, defined as average firing rate during times the animal was not in any trigger zone. Thus, the training set consisted of five expected firing rates: firing rate after zone-entry (1) at banana, (2) at cherry, (3) at chocolate, and (4) at nonflavored, plus a fifth control of expected firing rate (5) on the rest of the maze. Because of the inclusion of the fifth (average firing rate when not in any trigger zone) condition, the normalization factor is 0.20.

Calculating representations of previous, current, next opposite

In order to average across passes between different rewards, we first calculated the posterior probability for a given question (e.g. *p(Reward)* or *p(Zone)*) separately for each restaurant or zone. We then rotated the results based on the zone/reward in question to define a *current* zone/reward (the one the rat is currently encountering), a *previous* zone/reward (the one the rat had just left), a *next* zone/reward (the one the rat would encounter next), and an *opposite* zone/reward.

By utilizing ensemble decoding, we can effectively ask what recorded neurons are representing with the highest probability, taking into account both increases and decreases in firing rate. As shown in Supplemental Figures S8 and S9, the ensemble reliably differentiated entries into the different zones as well as the different rewards. As shown in Fig 2 from the main text, during normal behavior, the ensemble reliably represented the current zone on entry into it and the current reward on receipt of it.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Coricelli G, Dolan RJ, Sirigu A. Brain, emotion and decision making: the paradigmatic example of regret. Trends Cogn Sci. 2007; 11:258–265. [PubMed: 17475537]

2. Camille N, et al. The involvement of the orbitofrontal cortex in the experience of regret. Science. 2004; 304:1167–1170. [PubMed: 15155951]

3. Gilovich T, Medvec VH. The experience of regret: what, when, and why. Psychological Review. 1995; 102:379. [PubMed: 7740094]

4. Bell D. Regret in decision making under uncertainty. Operations Research. 1982:961–981.

5. Landman J, Manis JD. What might have been: Counterfactual thought concerning personal decisions. British Journal of Psychology. 1992; 83:473–477.

6. Loomes G, Sugden R. Regret theory: An alternative theory of rational choice under uncertainty. Economic Journal. 1982; 92:805–824.

7. Loomes G, Sugden R. Disappointment and dynamic consistency in choice under uncertainty. Review of Economic Studies. 1986; 53:271–282.

8. Bell D. Disappointment in decision making under uncertainty. Operations Research. 1985; 33:1–27.

9. Landman J. Regret: A theoretical and conceptual analysis. Journal for the Theory of Social Behaviour. 1987; 17:135–160.

10. Lee D. Neural basis of quasi-rational decision making. Curr Opin Neurobiol. 2006; 16:191–198. [PubMed: 16531040]

11. Coricelli G, et al. Regret and its avoidance: a neuroimaging study of choice behavior. Nat Neurosci. 2005; 8:1255–1262. [PubMed: 16116457]

12. Schoenbaum G, Chiba AA, Gallagher M. Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. Nat Neurosci. 1998; 1:155–159. [PubMed: 10195132]

13. Tremblay L, Schultz W. Relative reward preference in primate orbitofrontal cortex. Nature. 1999; 398:704–708. [PubMed: 10227292]

14. Padoa-Schioppa C, Assad JA. Neurons in the orbitofrontal cortex encode economic value. Nature. 2006; 441:223–226. [PubMed: 16633341]

15. Jones JL, et al. Orbitofrontal cortex supports behavior and learning using inferred but not cached values. Science. 2012; 338:953–956. [PubMed: 23162000]

16. Takahashi YK, et al. Neural estimates of imagined outcomes in the orbitofrontal cortex drive behavior and learning. Neuron. 2013; 80:507–518. [PubMed: 24139047]

17. Wilson RC, Takahashi YK, Schoenbaum G, Niv Y. Orbitofrontal cortex as a cognitive map of task space. Neuron. 2014; 81:267–279. [PubMed: 24462094]

18. Sul JH, Kim H, Huh N, Lee D, Jung MW. Distinct roles of rodent orbitofrontal and medial prefrontal cortex in decision making. Neuron. 2010; 66:449–460. [PubMed: 20471357]

19. Abe H, Lee D. Distributed coding of actual and hypothetical outcomes in the orbital and dorsolateral prefrontal cortex. Neuron. 2011; 70:731–741. [PubMed: 21609828]

20. Schoenbaum G, Nugent LS, Saddoris MP, Setlow B. Orbitofrontal lesions in rats impair reversal but not acquisition of go, no-go odor discriminations. Neuroreport. 2002; 13:885–890. [PubMed: 11997707]

21. Fellows LK, Farah MJ. Ventromedial frontal cortex mediates affective shifting in humans: evidence from a reversal learning paradigm. Brain. 2003; 126:1830–1837. [PubMed: 12821528]

22. Rudebeck PH, Saunders RC, Prescott AT, Chau LS, Murray EA. Prefrontal mechanisms of behavioral flexibility, emotion regulation and value updating. Nat Neurosci. 2013; 16:1140–1145. [PubMed: 23792944]

23. McDannald MA, Lucantonio F, Burke KA, Niv Y, Schoenbaum G. Ventral striatum and orbitofrontal cortex are both required for model-based, but not model-free, reinforcement learning. J Neurosci. 2011; 31:2700–2705. [PubMed: 21325538]

24. McDannald MA, et al. Model-based learning and the contribution of the orbitofrontal cortex to the model-free world. Eur J Neurosci. 2012; 35:991–996. [PubMed: 22487030]

25. Steiner AP, Redish AD. The road not taken: neural correlates of decision making in orbitofrontal cortex. Front Neurosci. 2012; 6:131. [PubMed: 22973189]

26. Cromwell HC, Schultz W. Effects of expectations for different reward magnitudes on neuronal activity in primate striatum. J Neurophysiol. 2003; 89:2823–2838. [PubMed: 12611937]

27. O'Doherty J, et al. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science. 2004; 304:452–454. [PubMed: 15087550]

28. Roesch MR, Singh T, Brown PL, Mullins SE, Schoenbaum G. Ventral striatal neurons encode the value of the chosen action in rats deciding between differently delayed or sized rewards. J Neurosci. 2009; 29:13365–13376. [PubMed: 19846724]

29. van der Meer M, Redish A. Covert expectation-of-reward in rat ventral striatum at decision points. Front Integr Neurosci. 2009; 3:1–15. [PubMed: 19225578]

30. Lavoie AM, Mizumori SJ. Spatial, movement- and reward-sensitive discharge by medial ventral striatum neurons of rats. Brain Res. 1994; 638:157–168. [PubMed: 8199856]

31. Setlow B, Schoenbaum G, Gallagher M. Neural encoding in ventral striatum during olfactory discrimination learning. Neuron. 2003; 38:625–636. [PubMed: 12765613]

32. Takahashi YK, et al. The orbitofrontal cortex and ventral tegmental area are necessary for learning from unexpected outcomes. Neuron. 2009; 62:269–280. [PubMed: 19409271]

33. Takahashi YK, et al. Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex. Nat Neurosci. 2011; 14:1590–1597. [PubMed: 22037501]

34. 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. 1995; 74:733–750. [PubMed: 7472378]

35. Roitman MF, Wheeler RA, Carelli RM. Nucleus accumbens neurons are innately tuned for rewarding and aversive taste stimuli, encode their predictors, and are linked to motor output. Neuron. 2005; 45:587–597. [PubMed: 15721244]

36. Landman, J. Regret: The persistence of the possible. Oxford University Press; 1993.

37. Connolly T, Butler D. Regret in economic and psychological theories of choice. Journal of Behavioral Decision Making. 2006; 19:139–154.

38. Aronson E. The effect of effort on the attractiveness of rewarded and unrewarded stimuli. Journal of abnormal and social psychology. 1961; 63:375–380. [PubMed: 13862515]

39. Arkes HR, Ayton P. The sunk cost and Concorde effects: are humans less rational than lower animals? Psychological Bulletin. 1999; 125:591.

40. Hare TA, Camerer CF, Rangel A. Self-control in decision-making involves modulation of the vmPFC valuation system. Science. 2009; 324:646–648. [PubMed: 19407204]

41. Krajbich I, Armel C, Rangel A. Visual fixations and the computation and comparison of value in simple choice. Nat Neurosci. 2010; 13:1292–1298. [PubMed: 20835253]

42. Bray S, Shimojo S, O'Doherty JP. Human medial orbitofrontal cortex is recruited during experience of imagined and real rewards. J Neurophysiol. 2010; 103:2506–2512. [PubMed: 20200121]

43. Walton ME, Behrens TE, Buckley MJ, Rudebeck PH, Rushworth MF. Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. Neuron. 2010; 65:927–939. [PubMed: 20346766]

44. Noonan MP, et al. Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. Proc Natl Acad Sci U S A. 2010; 107:20547–20552. [PubMed: 21059901]

45. Rushworth MF, Noonan MP, Boorman ED, Walton ME, Behrens TE. Frontal cortex and reward-guided learning and decision-making. Neuron. 2011; 70:1054–1069. [PubMed: 21689594]

46. Carmichael ST, Price JL. Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. J Comp Neurol. 1995; 363:615–641. [PubMed: 8847421]

47. Carmichael ST, Price JL. Sensory and premotor connections of the orbital and medial prefrontal cortex of macaque monkeys. J Comp Neurol. 1995; 363:642–664. [PubMed: 8847422]

48. Ongur D, Price JL. The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. Cerebral cortex. 2000; 10:206–219. [PubMed: 10731217]

49. Schilman EA, Uylings HB, Galis-de Graaf Y, Joel D, Groenewegen HJ. The orbital cortex in rats topographically projects to central parts of the caudate-putamen complex. Neuroscience letters. 2008; 432:40–45. [PubMed: 18248891]

50. Mailly P, Aliane V, Groenewegen HJ, Haber SN, Deniau JM. The rat prefrontostriatal system analyzed in 3D: evidence for multiple interacting functional units. J Neurosci. 2013; 33:5718–5727. [PubMed: 23536085]

51. Janabi-Sharifi F, Hayward V, Chen CSJ. Discrete-time adaptive windowing for velocity estimation. Ieee T Contr Syst T. 2000; 8:1003–1009.

52. Hart WE, Goldbaum M, Cote B, Kube P, Nelson MR. Measurement and classification of retinal vascular tortuosity. Int J Med Inform. 1999; 53:239–252. [PubMed: 10193892]

53. Zhang K, Ginzburg I, McNaughton BL, Sejnowski TJ. Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. J Neurophysiol. 1998; 79:1017–1044. [PubMed: 9463459]
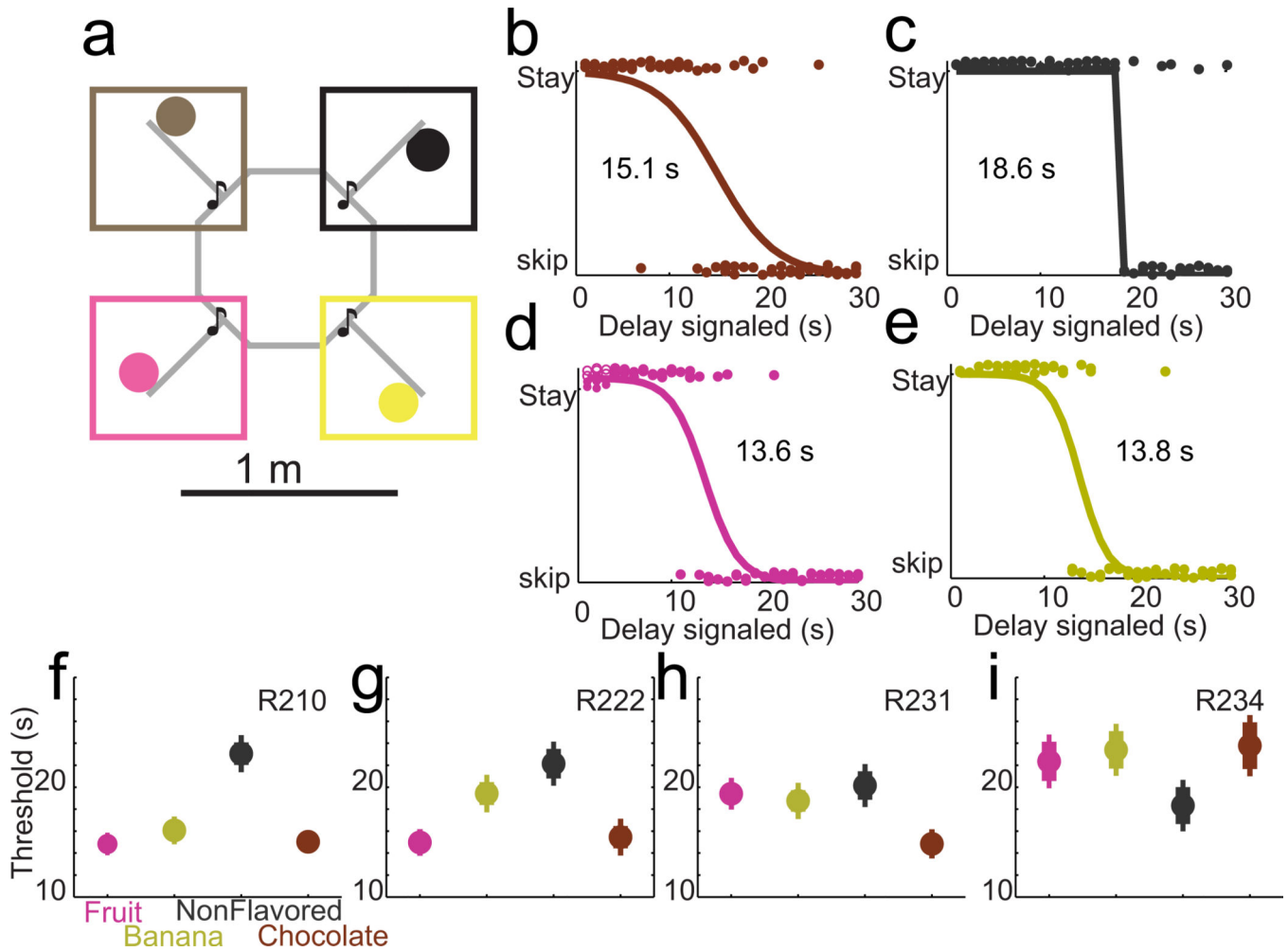
Figure 1. Restaurant Row and revealed preferences in rats

a, The Restaurant Row task consisted of a central ring with four connected spokes leading to individual food flavors. Rats ran around the ring, encountering the four invisible zones (square boxes) sequentially. Color reflects flavor: magenta=cherry, yellow=banana, black= nonflavored/plain, brown=chocolate. **b–e,** Rats typically waited through short delays but skipped long delays. Each panel shows the stay/go decisions for all encounters of a single rat running a single session (R210-2011-02-02). A small vertical jitter has been added for display purposes. Thresholds were fit as described in Methods. **f–i,** Each rat demonstrated a different revealed preference that was consistent within rat across all sessions, but differed between rats. Thresholds were fit for each flavor for each session. Each panel shows the mean fit threshold for a given rat, with standard error shown over sessions. An important consideration is to control for the possibility that rats were waiting for a specific cue before leaving the zone. The fact that rats either stayed through the entire delay or left after a very stable 3 seconds implies that rats were not waiting for a specific delay cue, but were making economic decisions based on the delay offered. (See Supplemental Fig S1).
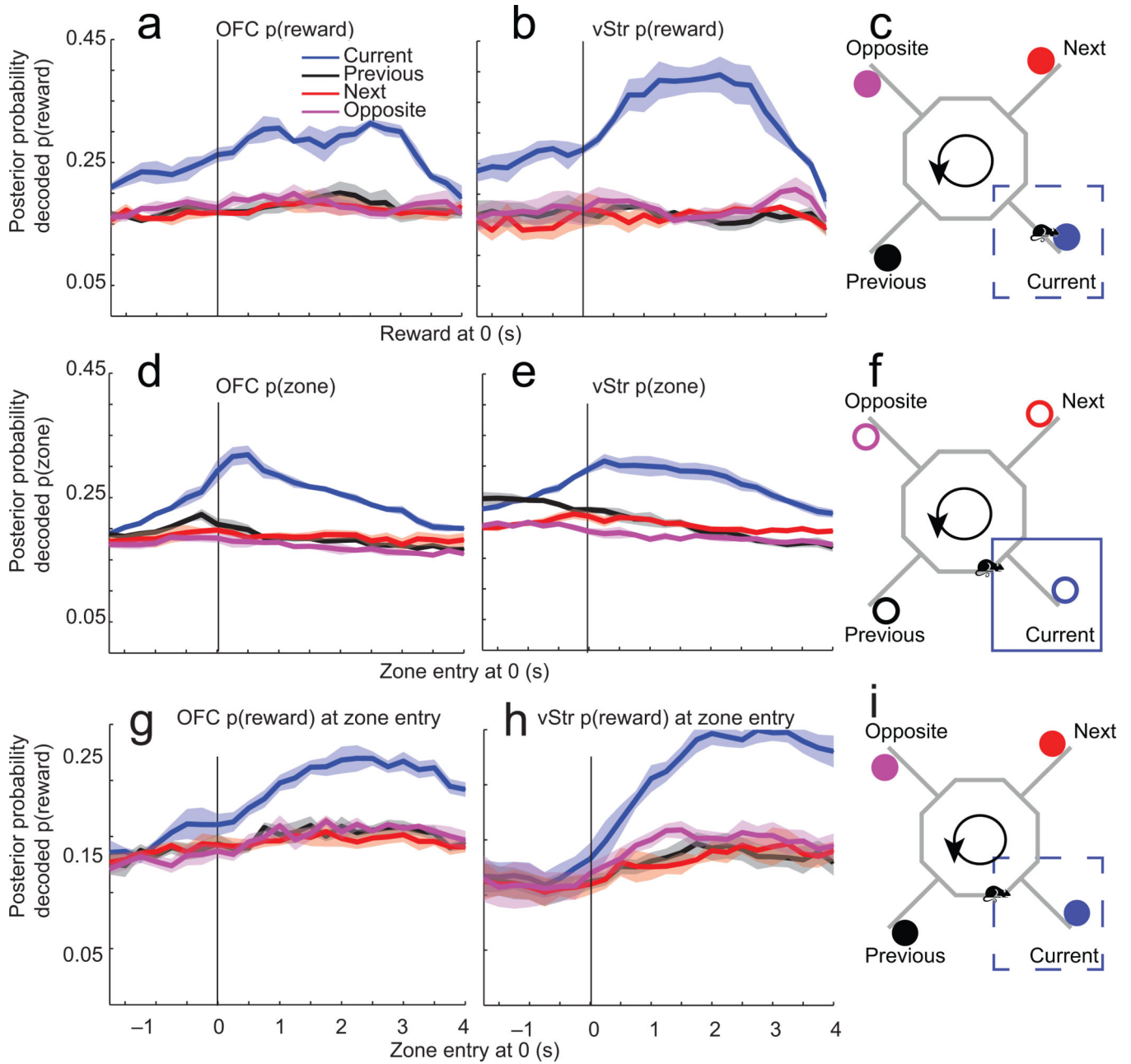
Figure 2. Ensembles in OFC and vStr represent the current reward and the current zone
a–b, *p(Reward) at Reward*, Defining the training set for decoding as activity at reward delivery and the test set as activity at each moment surrounding reward-delivery (shaded area represents standard error), the neural ensemble decoded the current reward reliably (distribution of current reward was determined to be significantly different, empirical cumulative distribution function, alpha = 0.05). *p(Reward)* is the posterior probability indicating representation of a given reward flavor as calculated by Bayesian decoding. **c**, Cartoon indicating that the training set is the reward types, and the test set is activity when the rat receives reward. **d–e**, *p(Zone) at Zone*, Defining the training set for decoding as neuronal activity at zone entry and the test set as neuronal activity at each moment

surrounding zone-entry, the neural ensemble decoded the current zone reliably. *p(Zone)* is the posterior probability indicating representation of a given zone entry as calculated by Bayesian decoding. **f**, Cartoon indicating that the training set is zone entry, and the test set is neuronal activity when the rat enters the zone, triggering the cue that signals the delay. **g–h**, *p(Reward) at Zone*, Defining the training set for decoding as neuronal activity at reward-delivery and the test set as neuronal activity at each moment surrounding zone entry, the neural ensemble at time of zone entry decoded the current reward type reliably. **i**, Cartoon indicating that the training set is the reward flavor, and the test set is neuronal activity when the rat enters the zone, triggering the cue/tone.
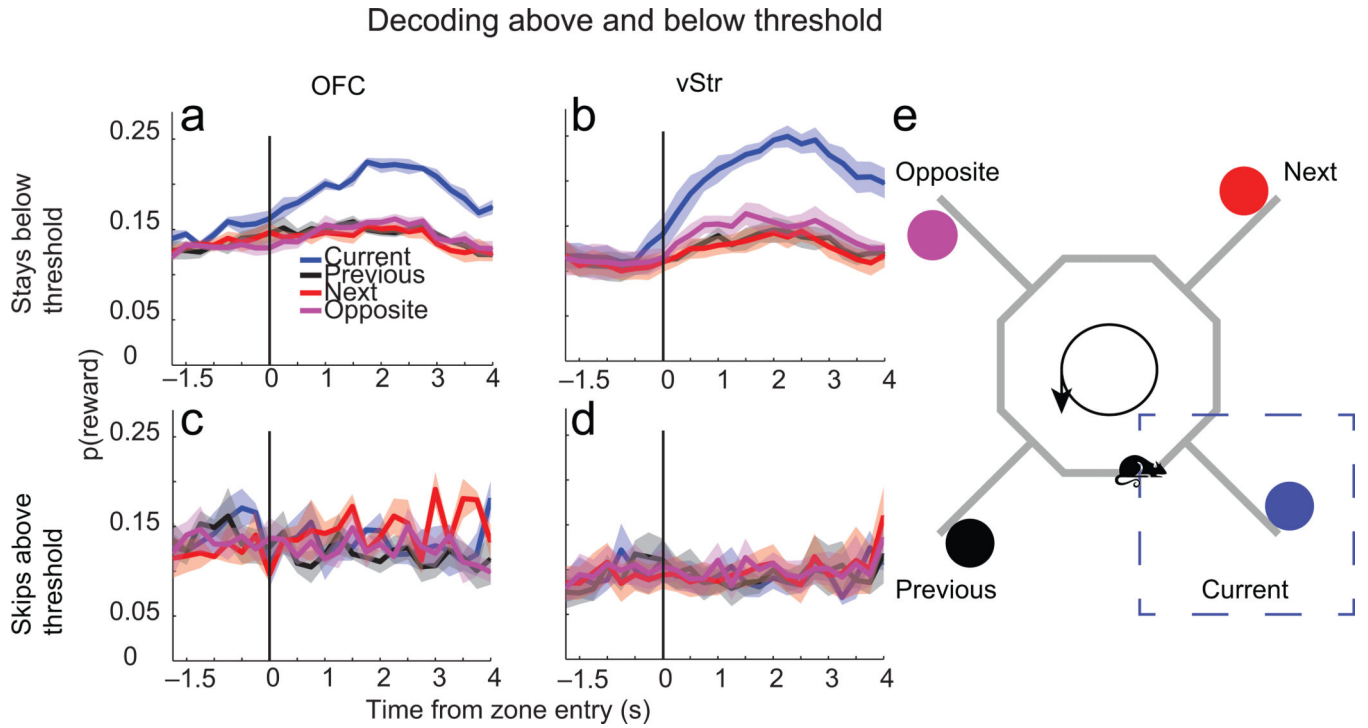
Figure 3. Representations of expected reward as a function of delay and threshold

In order to determine whether orbitofrontal (OFC) and ventral striatal (vStr) signals predicted behavior at time of zone-entry, we measured *p(Reward) at Zone* for all offers above and below the threshold for a given rat for a given flavor-reward-site (shaded area represents standard error). **a,b**, Low-cost offers in which the rat waited through the delay (distribution of current reward was determined to be significantly different, empirical cumulative distribution function, alpha = 0.05). **c,d**, High-cost offers in which the rat skipped out and did not wait through the full delay. **a,c**, OFC. **b,d**, vStr. **e**, Cartoon indicating that this decoding operation was based on a training set at the reward, but a test-set at zone-entry.
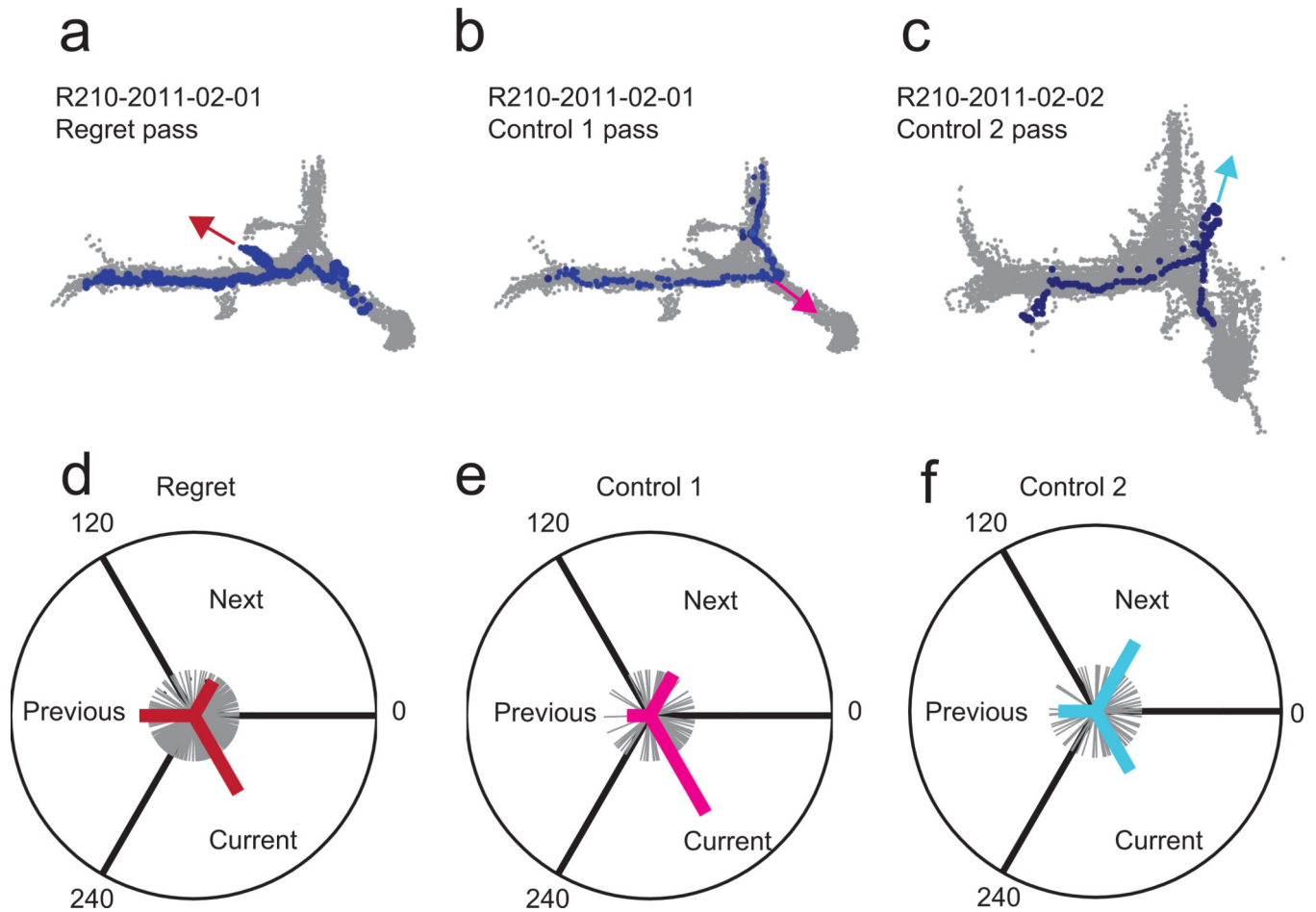
Figure 4. Behavioral responses in regret-inducing and control situations
All passes were rotated so as to align on entry into a "current" zone. Orientation was measured using the curvature measure as per Methods. **a–c**, examples of approaches for each of the three conditions: regret-inducing, control 1 (same sequence, took previous option), and control 2 (two long delays in a row). **a**, In a regret-inducing example, when the animal entered the zone, he paused and looked backwards towards the previous zone. **b,** In a control 1 example, the animal looked towards the current reward spoke, but proceeded on to the next zone. **c**, In a control 2 example, the animal looked towards the next zone, but turned back towards the current reward. **d–f**, Summary statistics. The first re-orientation event was measured as per Methods. Grey traces show all pausing re-orientations over all instances within that condition. Heavy line shows vector average within each 120 degree arc. **d**, In the regret-inducing conditions, rats tended to orient towards the previous zone or current spoke. **e**, In the control 1 conditions, rats tended to orient only towards the current spoke. **f**, In the control 2 conditions, rats tended to orient towards the next zone. The distributions in **d**, **e**, and **f**, were significantly different from each other (Watson's Circular U, see text).
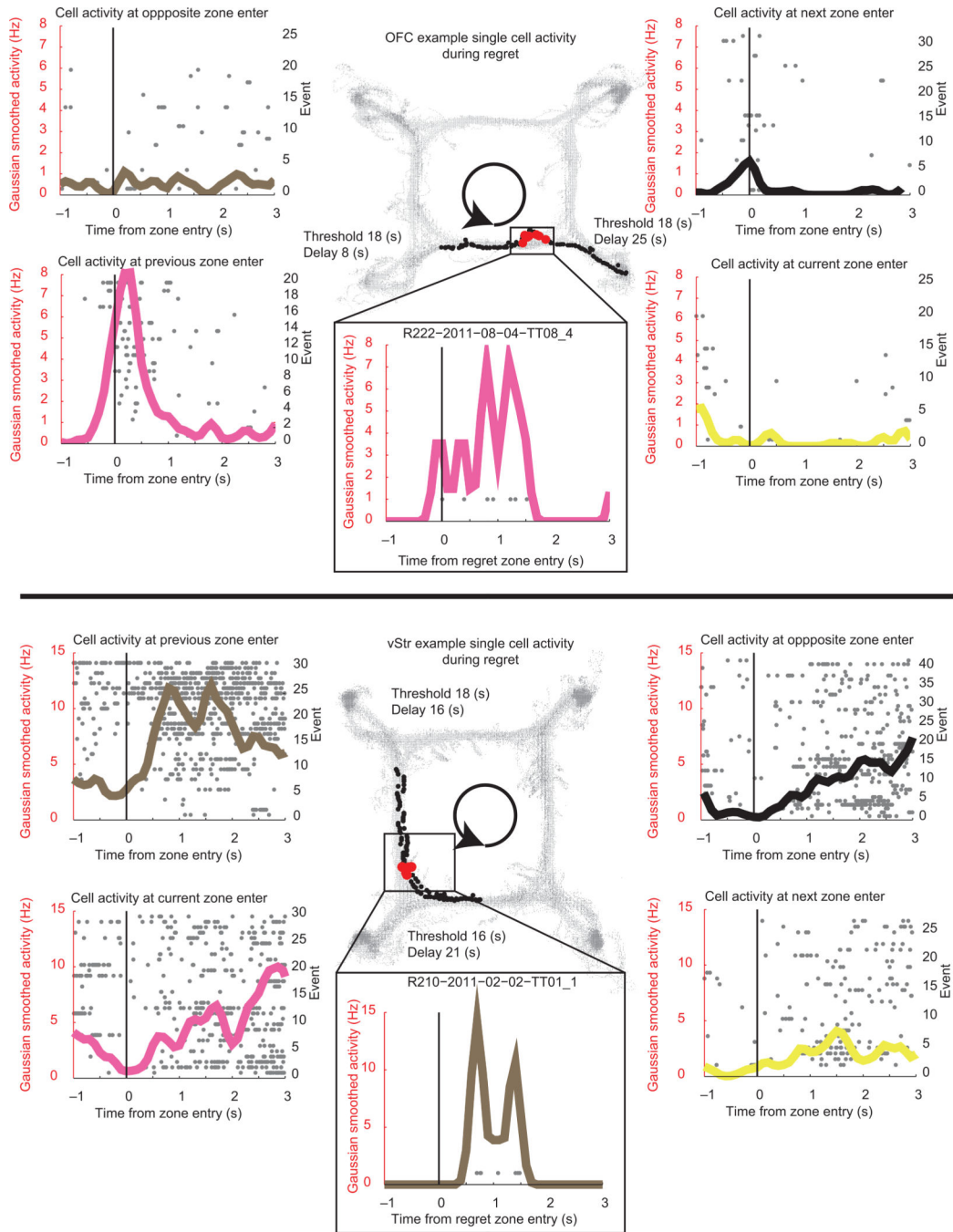
Figure 5. Single reward cells in OFC and vStr during regret-inducing situations
Top Panel. OFC Example Cell during regret-inducing situation. Grey dots represent individual spikes. Solid colored lines indicate Gaussian smoothed activity, sigma = 50ms. Black = nonflavored pellets, magenta = cherry flavored, yellow = banana flavored, brown = chocolate flavored. Black dots in the center panel represent behavioral samples during this particular instance. Red dots show spikes aligned to behavior. The rat traveled in a counterclockwise direction. The maze has been aligned so that the current zone is represented by the bottom right zone. This particular cell responded most to entry into the

cherry reward zone, and little to the banana reward zone. When the rat skipped a low cost cherry zone opportunity and encountered a high cost banana zone opportunity, the rat looked back towards the previous reward; and the activity of the cell approximated that of the cherry-zone-entry response. **Bottom Panel**. Display same as top panel, vStr example cell during a regret-inducing situation from the chocolate-reward zone to the cherry-reward zone.
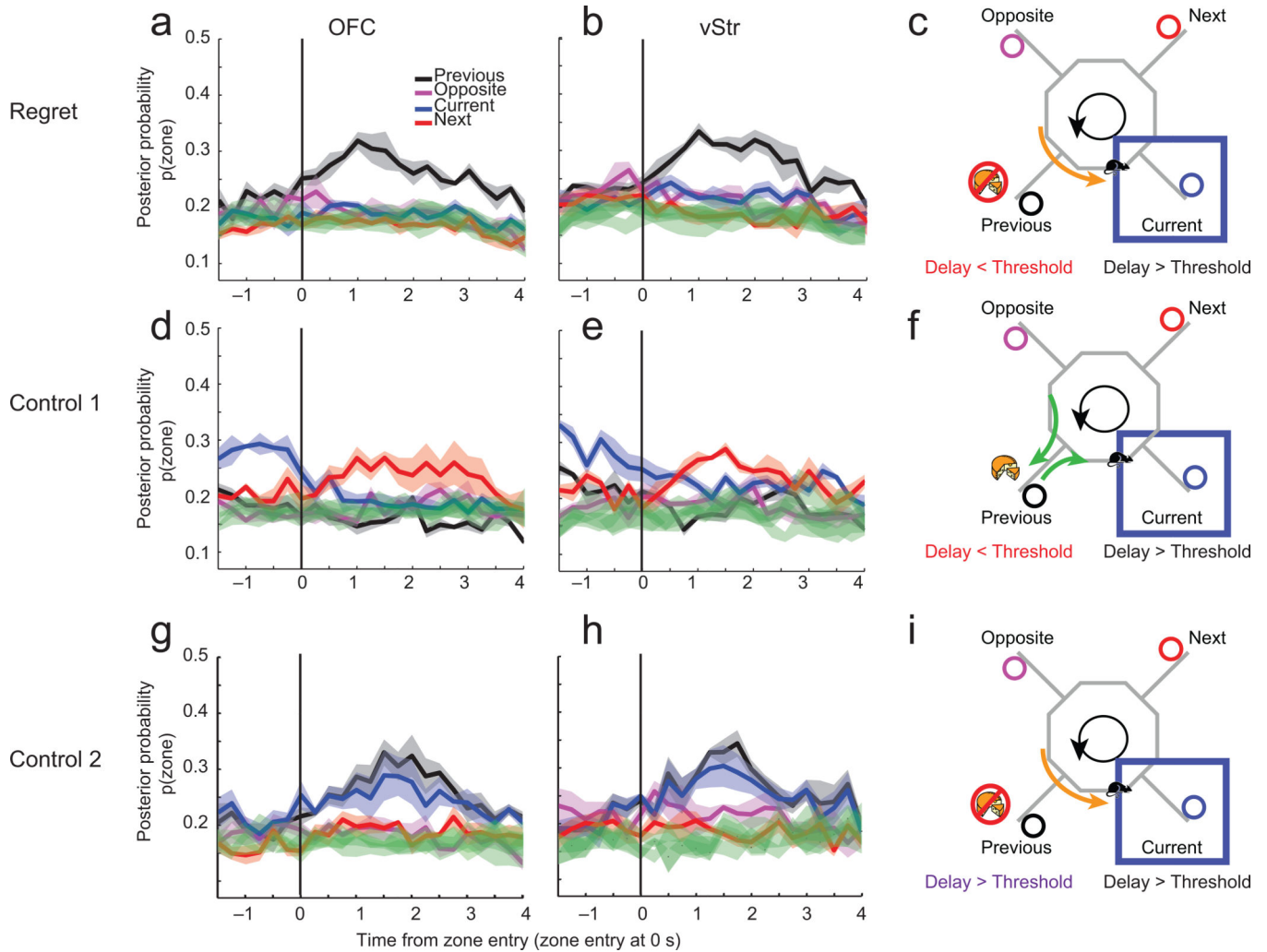
Figure 6. Neural representations in OFC and vStr represent the previous zone during behavioral regret instances

In regret-inducing conditions, the *p(Zone)* representation of the previous encounter was high after zone entry into the current zone for both OFC (**a**) and vStr (**b**) (shaded areas represent standard error). Green traces show decoding using shuffled ISIs. Decoding to the previous zone was significantly different from all other conditions, even after controlling for multiple comparisons (ANOVA, OFC: p << 0.001; vStr: p << 0.001, distribution significantly different as determined by empirical cumulative distribution function, alpha = 0.05). Panel **c** shows a cartoon of the conditions being decoded – the rat has skipped the previous offer, even though the delay was less than threshold for that restaurant, and has now encountered a delay greater than threshold for the current restaurant. In the control 1 condition, *p(Zone)* representation of the current zone increased until the rat heard the cue indicating a long delay, at which time, the representation changed to reflect the next zone. In control 1, *p(Zone)* representations to the current and next zones were significantly different from the other zones (ANOVA, vStr: p << 0.001; OFC: p << 0.001), although they were not different from each other after controlling for multiple comparisons (ANOVA, vStr: p = 0.074, OFC: p = 0.619). (**d**: OFC; **e**: vStr; **f**: Cartoon indicating condition.) In the control 2 condition,

p(Zone) representation of both the current and previous zones was increased when the rat heard the cue indicating a long delay (compared to other zones ANOVA, OFC: p << 0.001, vStr p << 0.001). (**g**: OFC; **h**: vStr; **i**: Cartoon indicating condition.) Decoding to the current and previous zones in control 2 were not significantly different from each other (ANOVA, OFC: p = 0.509; vStr: p = 0.268).
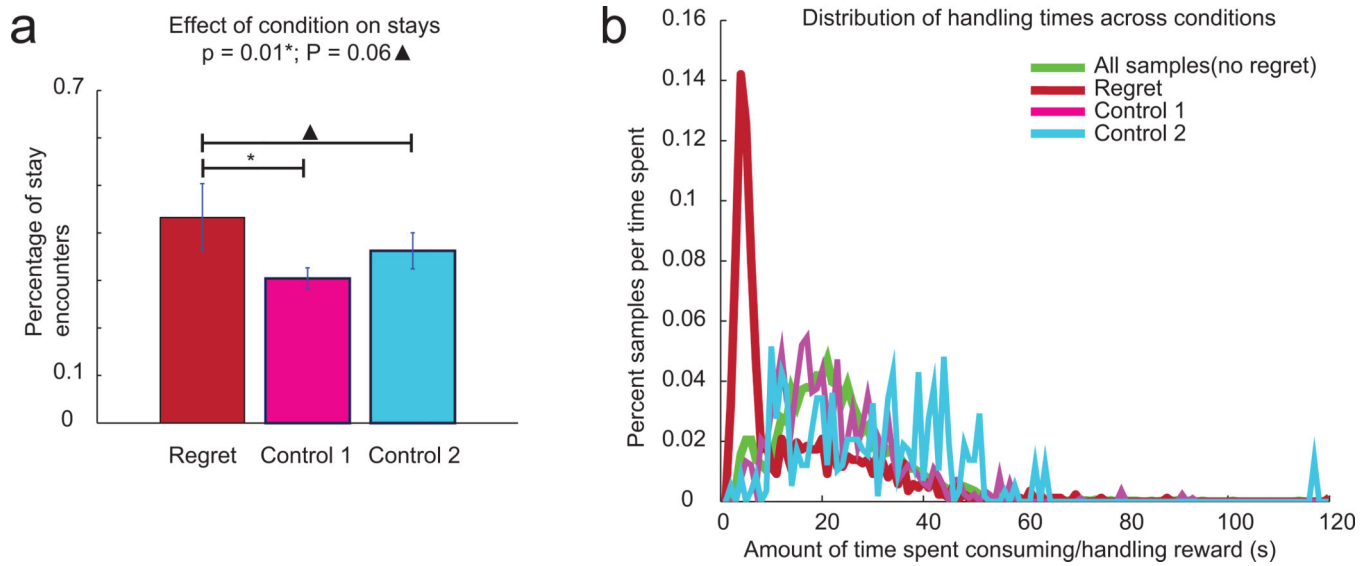
Figure 7. Behavioral changes following potential regret instances
a, Comparing the proportion of stays to skips during each condition revealed that rats were significantly more willing to wait for reward following regret-inducing instances compared to control 1 instances (Wilcoxon, p = 0.01) or control 2 instances (Wilcoxon, p = 0.06). **b,** Rats spent less time consuming reward during regret than during non-regret instances. Typical handling time mean = 25.3 s, standard deviation = 12.2 s, regret handling time mean = 15.2 s, standard deviation 14.2 s, control handling times are distributed the same as all non-regret handling times.
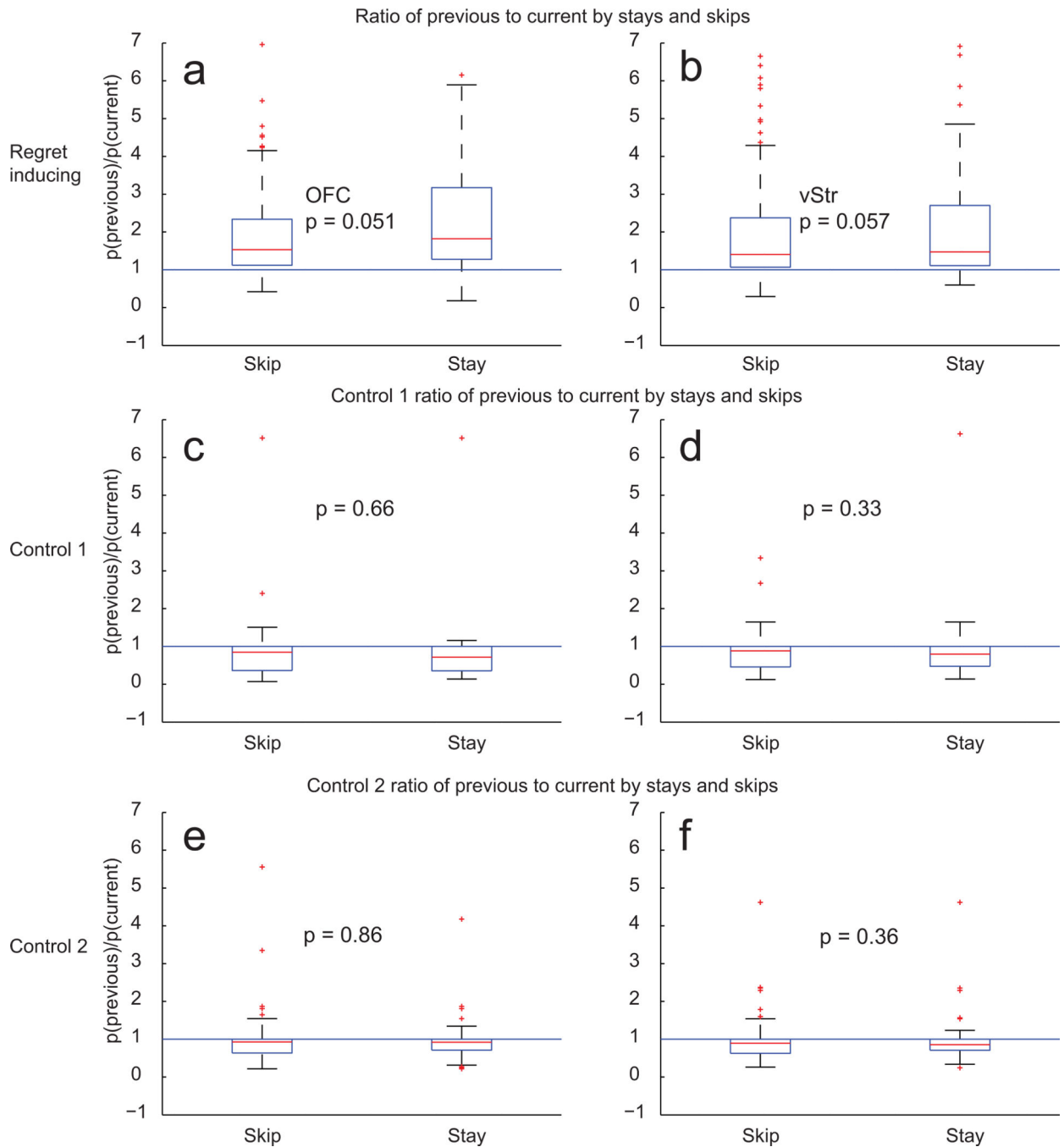
Figure 8. Behavioral and neurophysiological correspondences during regret
In order to determine whether the representations of previous reward were different when
the rat chose to stay at the high-delay (high-cost) current zone, we measured the ratio
between the *p(Zone)* representation of the previous zone against the *p(Zone)* representation
of the current zone from 0 to 3 seconds following zone entry for all conditions in the event
that the rat skipped or stayed. Each panel shows a box plot of the distribution of
p(Zone_previous)/p(Zone_current) ratios divided between stays and skips. Box limits are 25th and

75th percentiles, whiskers extend to data not considered outliers and outliers are plotted separately.

a, $p(Zone_{previous})/p(Zone_{current})$ ratios from OFC ensembles during regret-inducing conditions. **b**, $p(Zone_{previous})/p(Zone_{current})$ ratios from vStr ensembles during regret-inducing conditions. **d**, **e**, during control 1 conditions. **f**, **g**, during control 2 conditions. Following regret inducing instances, when rats were more willing to wait for reward, $p(Zone_{previous})$ was greater than $p(Zone_{current})$.