Representations of information value in mouse orbitofrontal cortex during information seeking

Jennifer J. Bussell[1,2], Ryan P. Badman[3,4], Christian D. Márton[5], Ethan S. Bromberg-Martin[6], L.F. Abbott[1,2], Kanaka Rajan[3,4], Richard Axel[1,2,7,*]

Author Affiliations:

[1] Department of Neuroscience, Columbia University, New York, NY, USA
[2] Mortimer B. Zuckerman Mind Brain Behavior Institute
[3] Department of Neurobiology, Harvard Medical School, Boston, MA, USA
[4] Kempner Institute, Harvard University, Boston, MA, USA
[5] Icahn School of Medicine at Mount Sinai, New York, NY, USA
[6] Department of Neuroscience, Washington University School of Medicine, St. Louis, MO, USA
[7] Howard Hughes Medical Institute
* Correspondence: ra27@columbia.edu

SUMMARY

Animals are motivated to acquire knowledge of their world. They seek information that does not influence reward outcomes suggesting that information has intrinsic value. We have asked whether mice value information and whether a representation of information value can be detected in mouse orbitofrontal cortex (OFC). We have developed an odor-based behavioral task in which mice choose to acquire information even though it does not alter the reward outcome. We observe that mice choose to acquire knowledge about uncertain reward and are willing to sacrifice water for information suggesting that knowledge is of intrinsic value to a mouse. We imaged neural activity in OFC while mice performed the information seeking task and observed different but overlapping populations of neurons responsive to odors predictive of information and odors predictive of water reward. Moreover, a nonlinear latent variable model recapitulated these distinct representations in the low-dimensional dynamics of OFC neuronal population activity. These data suggest that mice have evolved distinct pathways in OFC that represent the intrinsic value of knowledge and the extrinsic value of water reward. Thus, the desire to acquire knowledge is conserved in mice and the value of knowledge is represented in the OFC.

INTRODUCTION

Humans exhibit an innate drive to acquire knowledge. This knowledge may have intrinsic value and may also lead to the acquisition of rewards of extrinsic value. If the desire to know is innate, then it must have arisen by selection to enhance survival and reproduction. Thus, the evolution of an innate drive to acquire information of extrinsic value may ultimately have led to a quest for knowledge of purely intrinsic value. Early observations reveal that animals, like humans, acquire information about their world even if that information is of no apparent extrinsic value[1,2]. More recent experiments have addressed the nature of the brain representations of information involved in information seeking decisions[3–15], but the question as to how the brain identifies sensory stimuli as informative and computes their intrinsic value remains elusive.

Experimental efforts to understand the desire to acquire knowledge have largely involved paradigms in which subjects choose to receive information that resolves future uncertain outcomes[9,15–22]. When offered the opportunity to receive information that predicts a trial's reward outcome, animals often choose to acquire this knowledge despite the fact that the information does not alter the outcome of the task. Animals are even willing to sacrifice reward for information of no apparent extrinsic value[5,23–26]. These experiments suggest that the acquisition of information is of innate intrinsic value[27,28].

Experiments with monkeys have correlated neural activity with behavior in a paradigm in which animals are afforded the opportunity to choose information regarding the outcome of a task that offers liquid reward. Monkeys choose to acquire this predictive information despite the fact that it does not alter the reward outcome[4,15,23,28,29]. The task is complex as are the neural representations observed upon electrophysiological recordings. If both information, which is the internal reward, and juice, which is the external reward, are of value, each should contribute to a combined representation of value. Value is represented in multiple brain structures, but it is consistently observed in orbitofrontal cortex[30–37,37,38]. Individual neurons appear to respond to either information or external reward in the monkey orbitofrontal cortex (OFC)[23], suggesting the presence of distinct representations of intrinsic and extrinsic value. These

separate representations are consistent with a role for OFC in encoding distinct task variables of value that are integrated downstream to motivate a combined value-based choice[30,33,34,36,39–44].

In models of classical conditioning, the neurons responsive to the cues that predict both information and water must be reinforced to elicit observable value representations. Electrophysiological recordings reveal that midbrain dopamine neurons respond to both information and liquid reward in accord with classical mechanisms of reinforcement[28]. In functional MRI studies, similar information seeking tasks in humans elicit a BOLD signal in midbrain and prefrontal structures[5–7,14,45,46]. Efforts to identify a pathway that responds to information and ultimately activates these dopaminergic neurons have identified a cortex-basal ganglia circuit that extends from the anterior cingulate cortex (ACC) to the lateral habenula, a major regulator of dopaminergic cells[4,15,29]. Neurons in this pathway respond when reward outcomes are uncertain and diminish their response upon the acquisition of information that eliminates this uncertainty. Neurons in the lateral habenula send signals that reflect subjective value, integrating both the intrinsic reward of information and extrinsic reward[29]. This suggests a circuit in which information activates a cortex-basal ganglia loop, eliciting a dopamine-mediated reward prediction error. However, the nature of the brain structures capable of recognizing the innate value of knowledge and activating both the cortex-basal ganglia loop and an OFC population remain elusive.

We have asked whether mice value information and whether a representation of information value can be detected in mouse orbitofrontal cortex. We have developed an odor-based behavioral task in which mice are given a choice between informative and non-informative offers, each of no consequence to reward outcome. In this task mice learn the meaning of different odors and choose to acquire knowledge about uncertain reward despite the fact that this knowledge does not alter the reward outcome. Moreover, mice are willing to sacrifice water for information, suggesting that, in mice as in other organisms, knowledge is of intrinsic value.

In the mouse olfactory system, a neural representation of odor identity in piriform cortex is transformed into a representation of odor value in OFC by association with water reward[47]. Our task engages seven odors that predict the acquisition of information, provide information, and predict water reward. We recorded neural activity in OFC using a miniature microscope while mice performed the information seeking task. We observed different but overlapping populations of neurons responsive to odors predictive of information and odors predictive of water rewards. These two populations presumably represent the intrinsic value of information and the extrinsic value of water. The behaviors we observe reflect the integration of intrinsic and extrinsic value, and these representations in OFC must therefore be integrated to elicit the appropriate behavioral response. Thus, desire to acquire information is conserved in mice and the value of knowledge is represented in mouse OFC.

RESULTS

Mice value information

We have developed an information seeking task in which thirsty mice learn to enter a center port, where they receive one of two different odors that direct them either to the left or right side ports (forced trials) where they will receive a water reward on 25% of the trials (Fig1a). In one port, prior to the receipt of water the mice are exposed to one of two different odors that provide Information that reveals whether they will receive a reward on the current trial. In the other side port, the mice receive a different pair of odors that provide No Information as to the outcome of the trial. Whereas the reward probability is equivalent at the two side ports, the Information port differs in that it provides a mouse with information that predicts the reward outcome. Once the mice have learned the structure of the task a third odor is offered at the center port allowing a choice between the Information or No-Information port (choice trials). A bias for choice of the information port would imply that the mice seek knowledge of no apparent extrinsic value since this information does not alter the progress or the reward outcome of the task.

Mice learn the structure of this complex decision-based task in 5-8 weeks. When given a choice at the center port between visiting the Information or No Information side ports they choose the Information port on 68% (mean, sem 2%) of trials across several sessions of preference testing (Fig.1b). We performed multiple reversals of the side ports for each animal and observed that mice maintain their preference for the Information port despite its new location (Fig1b,c,FigS1b). Mice choose the Information port in 74% (mean, sem 7%) and 66% (mean, sem 10%) of the trials after the first and second reversal, respectively (FigS1b). Thus, mice prefer the Information port independent of the side on which it is located.

The majority of animals displayed a strong preference for information despite also displaying some degree of side bias (Fig1d). For example, we observed that four of 33 mice initially prefer the No Information port (FigS1a), but each of these mice fail to switch their preference on reversal sessions (Fig1b,d,FigS1b), indicating that their choice was dominated by side bias. Experiments were performed with seven female mice and 14 male mice, and we did not detect a sex difference in information preference (FigS1c). Thus, consistent with observations in other species, mice seek information about reward prediction even though this information does not alter reward outcome.

Two additional observations demonstrate that mice indeed learn the structure of the task and the information-predictive meaning of the odor cues at the center port. First, we measured the reaction time between the presentation of the center odor and entry into the side ports. Mice enter the Information port 250ms faster than they enter the No Information port on Forced trials (Fig1e, FigS1d). This finding is consistent with a preference for the Information port. We also recorded licks at the reward spout. Mice begin licking the water spout as soon as they enter the No Information port but withhold licks in the

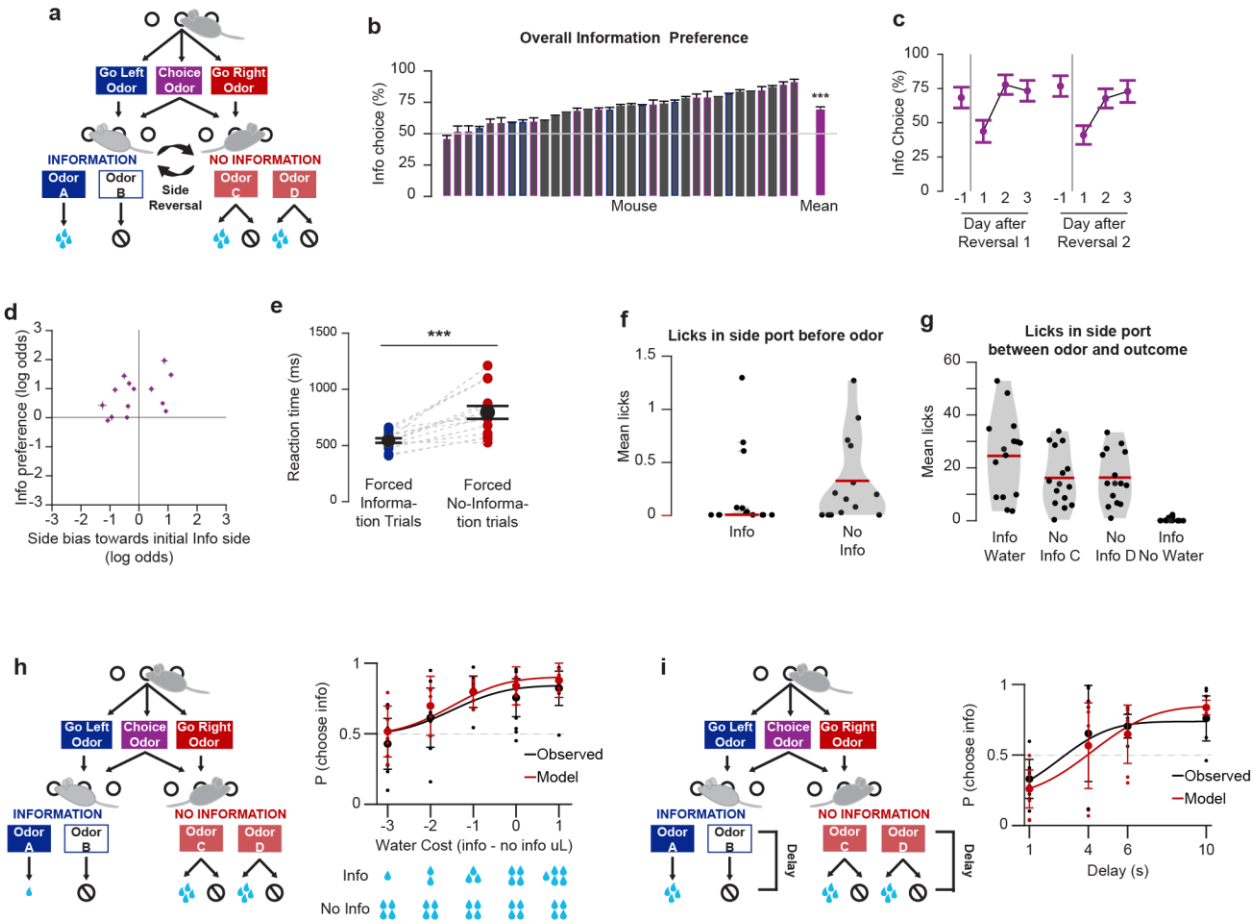

Figure 1: Mice seek information

(a) Information choice task diagram. Reward probability is 25% on each side, and odor C is provided on 75% of No Information trials, while odor D is provided on 25% of no information trials. Information is assigned to the left or right port at random across animals. (b) Overall information preference. Purple border indicates mice in initial cohort shown in c-g, blue border indicates imaged animals. Bars show mean preference for information port on choice trials across last three testing sessions prior to side reversal, +/- 95% confidence interval, p<0.001 sign-rank test. (c) Preference for information on choice trials across animals on sessions surrounding reversal of left-right side port information identities. In reversals, informative (AB) and non-informative (CD) odors and their reward contingencies are switched between the left and right side or vice versa. Center port odors remain the same, such that the same odor still directs the mouse to the left or right, but information content is changed. Points show mean +/- 95% confidence interval across animals. Data in (c-h) from initial cohort of 14 animals. (d) Influence of information preference vs. left-right side bias on choice. Plot shows coefficients for the logistic regression of current information side vs. left/right on each animal's choices across all preference testing days, including equal numbers of sessions with information on either the right or the left. Axes show log-odds of the variable's bias of choices. (e) Reaction times on forced information and forced no information trials. Blue and red points for each animal, black points show cross-animal mean +/- standard error, t-test for significance. (f) Licks at the information and No Information side ports prior to side odor delivery. Black points show mean licks per animal from side port entry until odor onset, red line indicates median. (g) Licks between side odor delivery and time of reward outcome. Black points show mean licks per animal from side port entry until odor onset, red line indicates median, for licks following the indicated side odor. (h) Influence of delay between side odor and reward outcome on choice preference. Task diagram indicates the manipulated delay period. Plot shows preference for information at different delay lengths. Points represent single animal mean preference across the last two sessions at each value, bars represent the cross-animal mean and 95% confidence interval. (i) Water and information value tradeoff preferences. Task diagram indicates water rewards on sessions with 1 drop on rewarded information trials and 4 drops on rewarded no information trials. Plot shows preference for information at different water amounts provided on 50% of trials on the information and no information side. Points represent single animal mean preference across the last two sessions at each value, bars represent the cross-animal mean and 95% confidence interval.

Information port until the informative odors are presented (Fig1f), indicating that they have learned that the Information side port odors reveal whether they will receive water on that trial. These observations indicate that the mice understand the predictive meanings of the 3 information-associated odors at the center port and distinguish the two ports by their information content.

We also observed behaviors in the side ports that reflect expected conditioned responses to the water reward value predicted by each odor cue. Following odor

A, the cue that predicted water at the Information port (100% reward), animals remained in the port and licked most frequently. Following the two odors at the No Information port (25% reward), mice remained in the port and licked less frequently. Following odor B, the cue signaling the absence of water at the Information port, the mice did not lick and left the port (0% reward; Fig1g; Fig S1e). We observed that mice leave the side port on nearly all unrewarded Information trials, which meant that their behavior differs between the Information and No Information ports during many of the trials. While it did not affect their access to water reward on these trials, leaving the Information port may be of value to a mouse, and this might contribute to their preference for the Information port. We therefore asked whether preference for information was strongly correlated with the frequency of leaving at the side port, which would be expected if leaving the port was a significant factor in information preference. We did not detect a strong correlation, either across animals or in each session (FigS1f,g). This suggests that animals primarily chose the Information port in order to obtain informative cues, not merely in order to express reward-conditioned responses after obtaining those cues.

The preference we observe for the Information port presumably reflects the value of information. If so, then mice should be willing to pay for information, a crucial signature of value in decision-based information seeking tasks[5,23–25,48]. We therefore asked whether mice were willing to sacrifice water for information by varying the amount of water reward on information trials over blocks of several sessions. Trials were performed in which the Information port provided from 1-6 drops of water, whereas the No Information port consistently provided 4 drops of water, and on both sides water was provided with 50% probability. Mice continue to prefer information even at the expense of diminished reward (Fig.1h). We fit a psychometric curve and decision model (below) to their preference as water was diminished at the Information port and calculated that mice assign a subjective value to information equivalent to 4-6uL, or more than half of their expected reward (Fig.1h, black). Mice are therefore willing to sacrifice water to receive information, suggesting that their choice is an integration of intrinsic and extrinsic value.

In our information seeking task, mice move from the center port to a side port where they experience a 10s delay before receiving a water reward. We expect that were we to shorten this delay, if mice value information because it grants them access to informative cues in advance of the delayed reward, the value of information should also be reduced because it either diminished the duration of pleasurable anticipation or aversive uncertainty[6,16,19,49]. We have therefore manipulated the subjective value of information in this task by varying the length of delay on both sides, Information and no information, across blocks of sessions. We observed that the preference for information decreases with decreasing delay, both within individual mice and across the population (Fig.1i, black). We fit a psychometric curve and decision model (below) to the relationship between delay and information preference and found that mice become indifferent to information at a delay of 1-4s across animals.

We developed a model to examine the behavioral effects of delay and water reward history on the choice to receive information under the hypothesis that choices in our task are guided by separate but interacting reinforcement learning-like computations of intrinsic information value and extrinsic water value. By manipulating either the reward delay or water amount and having the model update the values assigned to the ports based on experience, we fit a Rescorla-Wagner[50,51] model to the choices of mice during the above water tradeoff and delay tradeoff sessions. This model included two separate prediction error value functions, one for water and one for information, that capture the relationship between water reward and information choice history and information preference. We also fit a delay-related discount exponent for the value of information for each animal to quantify the degree to which delay modulates information preference (Fig.S2, Eq. 6). This model fit all mice to ~70-80% accuracy and captured preference shifts versus water trade-offs and delay lengths (Fig.1hi,i, red). We compared this model to reduced models in which key terms were removed and found that the full model with learning of both water and delay-discounted information values most accurately captured choice behavior (Fig.S2). This decision model provides additional evidence that mice display a preference for the Information port based on the subjective value of information that depends at least in part on the duration of the delay between information and reward.

A representation of information prediction in OFC

The behavioral preference for information implies that information is of subjective value to the mouse. In the mouse olfactory system, a neural representation of odor identity in piriform is transformed into a representation of extrinsic value in OFC[47]. We have therefore asked whether we can detect representations of both extrinsic and intrinsic value in OFC as mice perform our information seeking task. We recorded calcium signals reflecting neural activity in 1138 excitatory neurons in OFC using miniaturized microscopes, and we registered cells across sessions for 6-8 weeks (Fig2a,FigS3a,b). Neurons in mouse OFC represent task variables including stimulus identity, motor action, confidence and task context as well as reward value[31–35,40,41,52–54]. Given the complexity of our task we expect that OFC will contain multiple overlapping representations that reflect mixed selectivity in individual cells[55]. We initially examined OFC activity in response to odors presented at the center port where mice receive two odor cues that direct them to either the Information or No Information port or a third odor that permits a free choice between the two ports. If information is of intrinsic value, the center port odors that result in movement to the Information port, whether

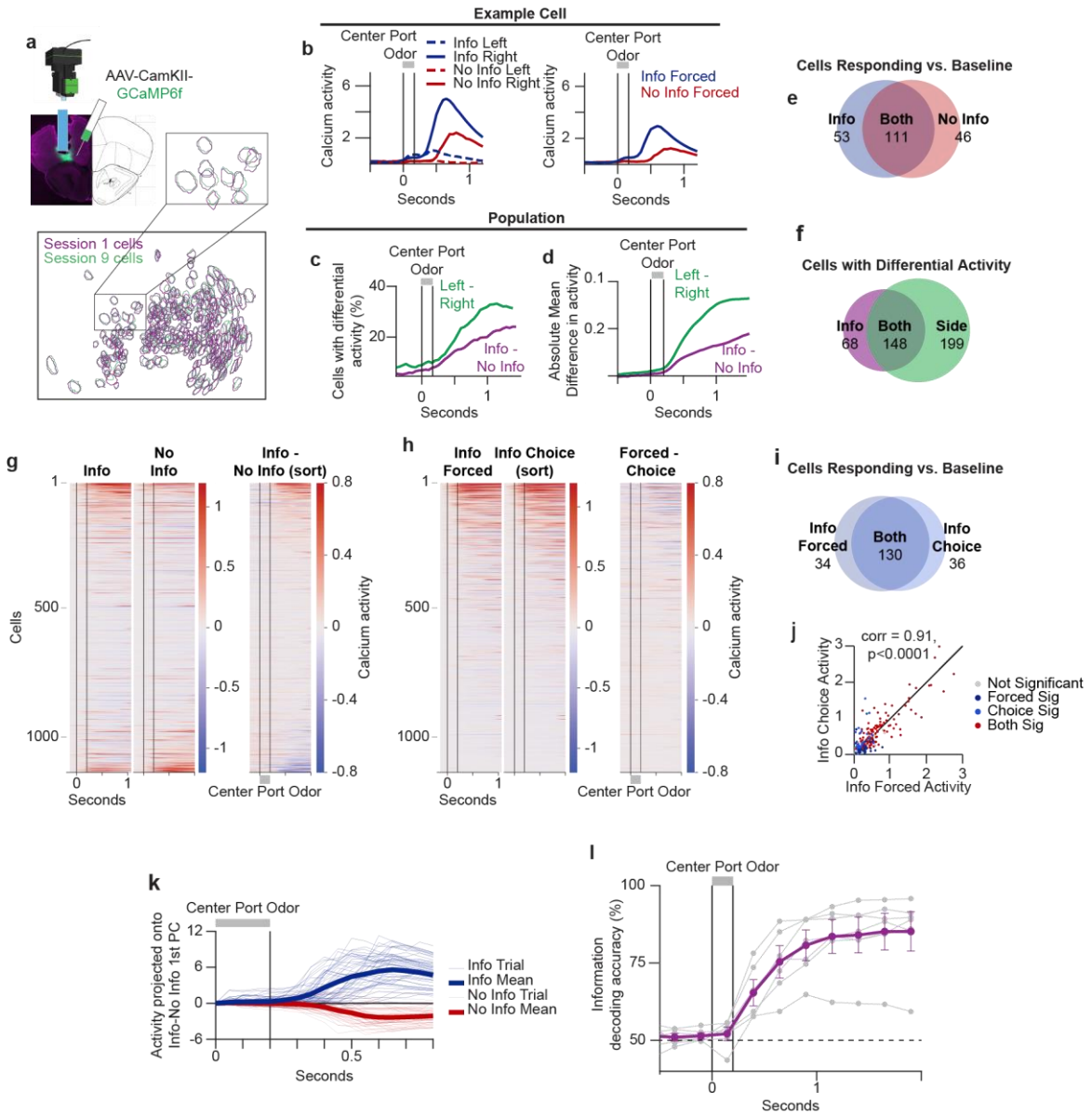

Figure 2: A representation of information prediction in OFC

(a) Schematic of miniature microscope imaging and GCaMP-encoding viral injection locations in OFC and schematic of cell registration. (b) Responses of a single example cell to center port odors. Left, mean activity response to information and No information forced trial odors when information is provided in the right and left side ports. Right, mean activity response to information and no information forced trial odors across both side ports. (c) Percent of cells with differential activity between trials in the indicated conditions at each frame surrounding center port odor presentation. (d) Mean absolute value of the difference in activity across cells between the indicated conditions. (e) Number of cells responding to the information forced and no information forced trial center port odors and their overlap. N=1138 total recorded cells. (f) Number of cells with differential activity between the information and no information forced trial center port odors and between odors forcing choice of the left and right side, and their overlap. (g) Population activity of responses to the Information and no information forced trial center port odors and their difference. Each row shows the mean-subtracted activity for each cell across all the panels. All plots are sorted by the difference between Information and no information responses. (h) Population activity heatmap of responses to the information forced trial odor and the choice center port odor on trials in which information was chosen and their difference. Plots as in (g), sorted by the response to the information choice odor. (i) Number of cells responding to the information forced and information choice trial center port odors and their overlap. (j) Correlation between activity in response to center port information forced trial odor and choice trial odor on information trials. Each point indicates the mean-subtracted activity in response to each odor for a single cell. (k) Projection of single trial and mean activity onto the first principal component of the difference between mean information and mean no information center port odor responses across cells. (l) Decoding information vs. no information trials from activity surrounding center port odor presentation across both forced and choice trials. Purple, mean +/-95% CI across mice, grey=individual animals.

forced or choice, should serve as conditioned stimuli predicting the acquisition of information.

We initially focused on the differential response to center port odors that direct the mice to Information or No Information side ports on forced trials across reversals of left and right side identities. This allowed us to disambiguate value representations from representations of side movement. One example cell reveals high activity in response to the odor predictive of information in the right port (Fig.2b) and responds more weakly to the odor predictive of information in the left port or to odor signaling no information. This cell therefore exhibits a preference for odor that predicts information at the right port. Across the population, 19% of the recorded OFC cells displayed a differential response to the odors predicting information or no information (mean Information – mean no information) ($p<0.05$, bootstrap shuffling, Fig.2c,g).

If these neurons indeed represent the value of odors predictive of information, a CS+, we would expect that this representation should also be activated in choice trials in which mice freely choose the Information side port. We observed that the odors that result in either the forced or chosen movement to the Information port activate largely overlapping representations despite the different odor identities and task context (Fig2h). 79% of the neurons that respond to odors that forced movement to the Information port are also activated in information trials in which the mouse chose the Information port (Fig2i). The center port odor response within each cell was strongly correlated between forced information and choice information trials (corr=0.91, $p<0.0001$, Fig2j). Moreover, center port odor responses were more strongly correlated between forced information and choice trials (FigS3c, corr=0.92) than forced no information and choice trials (FigS3d, corr=0.80), consistent with these responses reflecting the prediction of information.

We observed that the responses to both the information and no information center port odors reflect mixed selectivity. We registered cells across reversals of side port identity to distinguish the representation of the prediction of information and the prediction of side. We observed that 30% of cells exhibit a differential response to odors that direct movement either to the left or right port (Fig.2b-d). These cells show significant overlap with cells that exhibit a differential response to information. 68% of the neurons that represent information also encode side (Fig2f). The neural response to side across the population was large and accounted for 25% of the total variance. The population activity of neurons responsive to odors predictive of information was lower but also robust, accounting for 7% of the total population activity variance. This population of neurons predictive of information responds differentially to odors that predict information and odors that predict no information. We interpret this population response as a representation of the expectation of information and isolated it by balancing data across reversals.

We next asked whether this population of neurons represented the animal's expectation of information on a trial-by-trial basis. We calculated the principal components of the difference in each cell's mean response to odors predictive of information and no information on all trials (both forced and choice). This revealed a single mode that captured >80% of the variance in information-related activity. We then projected each trial's activity, and the cross-trial mean for each type, onto this information-no information axis. This analysis revealed trial-by-trial population activity for odors encoding information on both forced and choice trials (Fig.2k). We then used a linear classifier to decode information versus no information trials from center port odor-related activity on all trials (both forced and choice), as well as choice trials separately (Fig2l,FigS3e). These analyses demonstrate that OFC neurons encoded the prospect of receiving information on a trial-by-trial basis on both forced and choice trials. The center port odor predictive of information may therefore be considered a CS+ reflecting the animal's preference for information (Fig.2g).

We have considered the possibility that the center odor representation may also reflect the animals' movement to execute their choice since we observed that mice reacted faster on information trials. However, the representation of odors predictive of information is not a reflection of port entry, movement or speed. Within center port responses predictive of either information or no information, OFC population activity was only very weakly correlated with reaction time and thus the animal's speed of movement (Fig.S4a,b). Further, activity in OFC was more closely aligned to odor onset than either the animal's exit from the center port or arrival at the side port (Fig.S4c). These observations suggest that the representation of the expectation of information reflects the prediction of information and not a movement-related change in neural activity.

Representations of value upon the receipt of information

We have identified a representation of odor cues delivered at the center port predictive of information. If these cues represent a CS for information, it follows that the odors that provide information at the side port represent a US for information. We therefore examined responses to the side port odors that provide information as to the different possible reward outcomes (FigS5). The responses across the population to odors A and B at the Information port were highly correlated (correlation=0.73, $p<0.001$, Fig3i). We observed a subpopulation of cells that responded to both odor A (water) and odor B (no water) (Fig3e,f), consistent with a shared response to both of these informative odors. One example of these cells is shown in Figure 3b (top). These data suggest that odors A and B elicit a shared representation of the receipt of this information.

The interpretation of neural activity at the information port is complex, since odor A provides information and is also a CS+ predictive of subsequent water reward. In accord with this, we observed cells that respond more to odor A than odor B (Fig3g,i,FigS5d). An example cell (Fig3b bottom) responds most to odor A (100% reward), less to no information odors C and D (25% reward), and

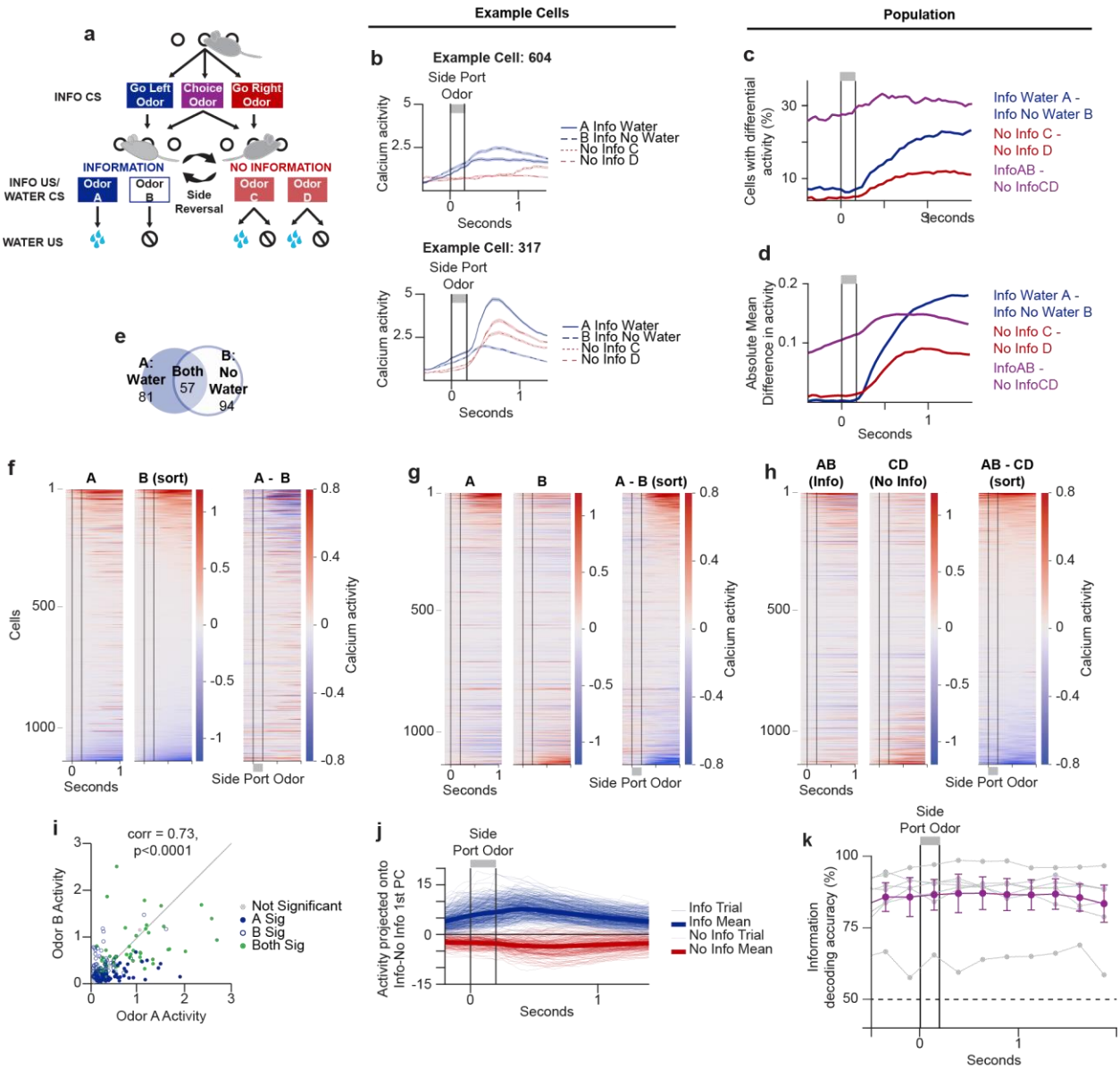

Figure 3: Representations of value upon the receipt of information

(a) Information choice task diagram. (b) Responses to odors presented at the side ports for two example cells. Top, this cell responds to both odor A and odor B, the two odors on the information side. Bottom, this cell responds to all four side port odors in proportion to the water reward they predict. (c) Percent of cells with differential activity between trials in the indicated conditions at each frame surrounding side port odor presentation. (d) Mean absolute value of the difference in activity across cells between the indicated conditions. (e) Number of cells responding to side port odors A and B and their overlap. N=1138 total recorded cells. (f) Population activity heatmap of responses to odor A and odor B and their difference. Each row shows the mean-subtracted activity for each cell. All columns are sorted by the response to odor B. (g) Population activity heatmap of responses to odor A and odor B and their difference. Plots as in (f) sorted by the differential response to odor A (right). (h) Population activity heatmap of responses to odors A and odor B, odors C and D, and their difference. Plots as in (f), sorted by the differential response (right). (i) Correlation between activity in response to side odor A and side odor B. Each point indicates the mean-subtracted activity in response to each odor for a single cell. (j) Projection of single trial and mean activity onto the first principal component of the difference between mean odors AB and mean odors CD side port odor responses across cells. (k) Decoding activity on information vs. no information trials surrounding side port odor presentation. Purple, mean +/-95% CI across mice, grey=individual animals.

least to odor B (0% reward) and may therefore encode the water value signaled by these odors. This response pattern is consistent with the suggestion that OFC encoding of the side port odors includes both representations of water value and representations of the receipt of information. We also observe a population of cells that exhibit mixed selectivity and respond both to the receipt of information and the prediction of water reward (FigS6d). Thus, at the Information port, odor A elicits representations of odor A as a US for information and a CS+ for water reward, whereas the odor B response may only reflect a representation of the US for information.

We next examined the mean response to odors A and B compared to odors C and D to identify a differential response specific to the receipt of information (Fig3h). We observed that 33% of OFC cells displayed differential responses to odors AB versus odors CD (Fig3c,d). We calculated the principal components of the difference in each cell's mean response to the AB versus CD odors. We then projected the activity along the information-no information axis defined by the first principal component vector. This analysis revealed trial-by-trial population activity representing the receipt of information (Fig.3j). This response to the receipt of information was also evident in the mean absolute difference in population activity between information and no information across the trial (FigS6a). The differential information-no information activity increased after the center port odor and continued to increase through the receipt of the side port odor, consistent with OFC representing both the prediction and receipt of information. The differential response to the receipt of information reflects two separate subpopulations, one that represents the receipt of information (AB>CD, FigS7j), and a second that represents the absence of information at the side port (CD>AB, FigS7k). Importantly, these responses do not result from the choice of the left or right side port, since they are also observed in reversal trials. We used a linear classifier to decode information versus no information trials from the activity of the entire OFC population (Fig3k), and this classifier's accuracy peaked at the receipt of the side port odors, consistent with the existence of a representation of the US response to the receipt of information.

At the No Information port, responses to odors C and D were highly correlated (corr=0.91, p<0.0001, FigS5b,c) as expected since neither provides information and both precede 25% water reward (FigS5f,g). A small fraction of OFC cells differentiated odor C from odor D (10%, Fig3c,d), and we attribute this response to the fact that odor D was presented on 25% of no information trials, while odor C was presented on 75% of no information trials, to control for the frequency of odors A and B (FigS5a). We observed a CD-responsive sub-population of cells that does not overlap with the AB-responsive subpopulation, does not respond to odor B, and responds only weakly to odor A (FigS6k). Instead, these CD cells appear to respond to the absence of information. Thus, we observed representations of both the information and no information sides as well as responses to information as a US and the CS prediction of water reward.

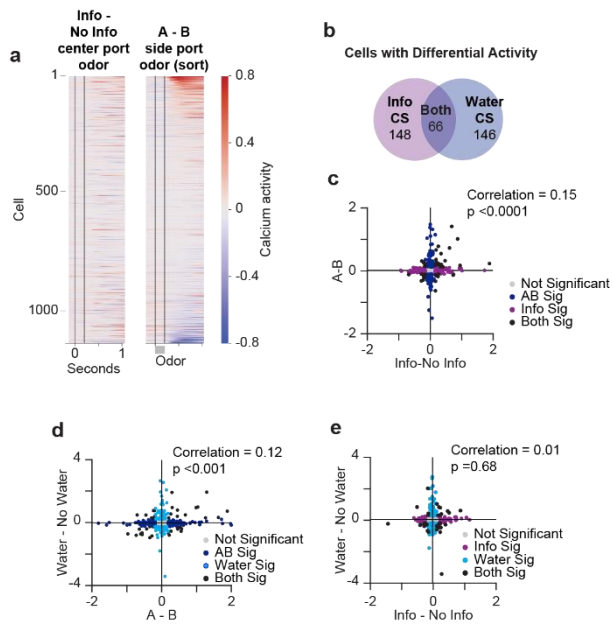

Figure 4: Distinct neural populations represent intrinsic and extrinsic value

(a) Comparison of information CS+ and water CS+ population activity. Left, difference between mean responses to the information and no information forced trial center port odors. Right, difference between mean responses to information side port odor A (water) and odor B (no water). Each row shows the mean-subtracted activity for each cell across both panels. Both plots are sorted by response to odor A-odor B. (b) Number of cells with differential activity between the information and no information forced trial center port odors and between the information side port odors A and B, and their overlap. N=1138 total recorded cells. (c) Correlation between the difference in activity in response to the information and no information forced trial center port odors (information CS+) and the difference in activity between the information side port odor A (water) and B (no water) (water CS+). (d) Correlation between the difference in activity in response to the side port odor A and odor B (water CS+) and the difference in activity between water reward and no water reward in the No Information port. (water US). (e) Correlation between the difference in activity in response to the information and no information forced trial center port odors (information CS+) and the difference in activity between water reward and no water reward in the No Information port. (water US).

Distinct neural populations represent intrinsic and extrinsic value

The neural activity in response to odor predictive of information at the center port may reflect a CS+ for the receipt of information, whereas the activity in response to odor predictive of water at the information side port reflects a CS+ representation for water reward. We asked whether the CS+ representation of information is distinct from the CS+ representation of water. We compared the population of cells exhibiting differential activity to odor predictive of information at the center port (CS+ info) with the population of neurons that show a differential response to odor that predicts water (CS+ water) at the side port (Fig4a). We observe that these two CS+ representations (information and water, respectively),

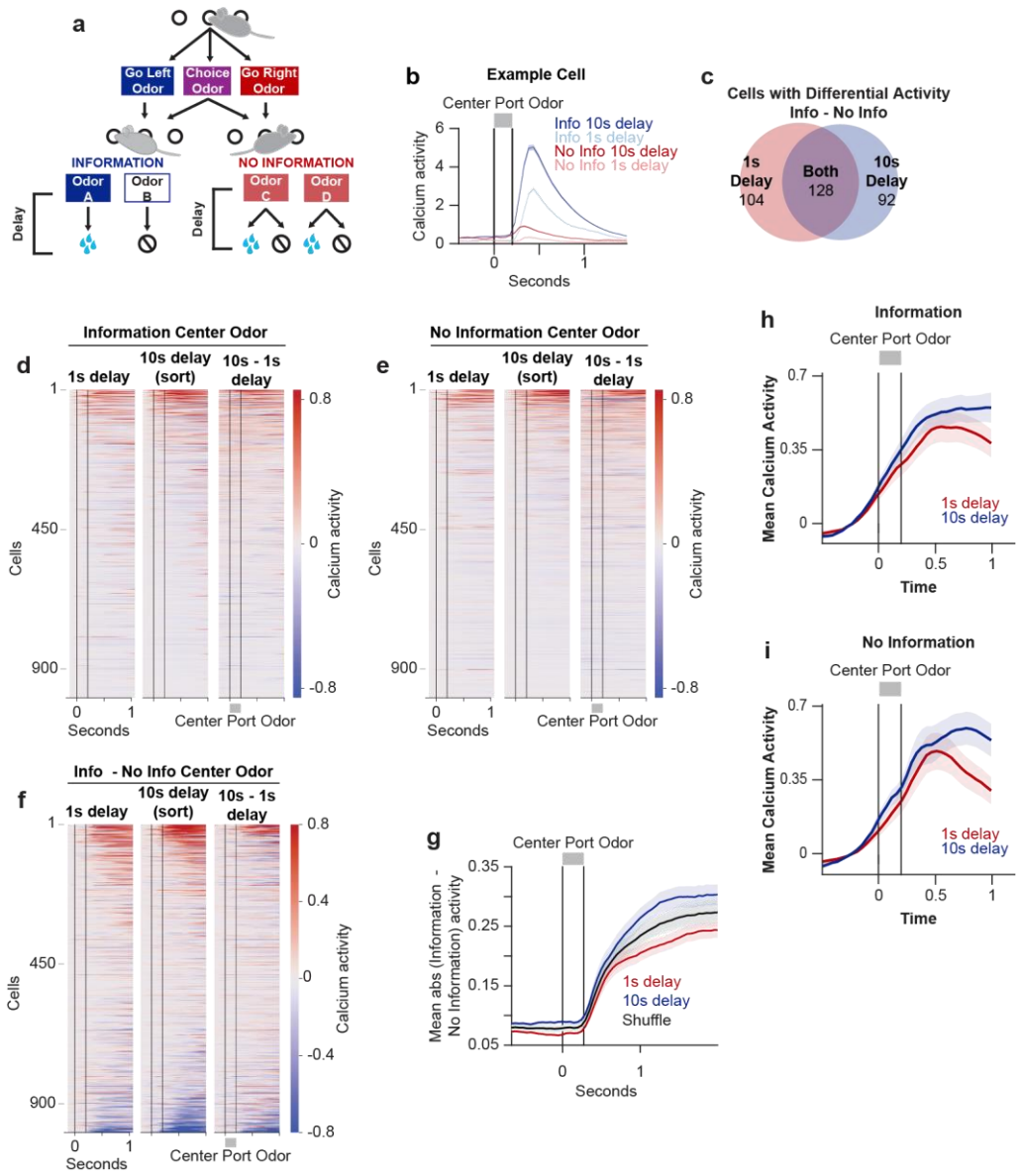

Figure 5: OFC signals reflect information value

(a) Information seeking task diagram illustrating the delay period that was manipulated. (b) Responses of a single example cell to information and no information forced trial center port odors in sessions with different delays. (c) Number of cells with differential activity between the information and no information forced trial center port odors in sessions with a 10s delay and sessions with a 1s, and their overlap. N=1138 total recorded cells. (d,e) Information and no information prediction population responses at 10s and 1s delay. Population activity of responses to the information or no information forced trial center port odor in sessions with 10s and 1s delay and their difference. Each row shows the mean-subtracted activity for each cell, or difference in mean activity, across all panels. All plots are sorted by the response on 10s delay sessions. (f) Population information CS+ representation at 10s and 1s delay. Population activity of the differences in responses to the information and no information forced trial center port odor in sessions with 10s and 1s delay. Right column shows the difference between the differences at 10s – differences at 1s. Each row shows the or difference in mean activity for each cell across all panels. All plots are sorted by the response difference on 10s delay sessions. (g) Mean information CS+ representation at 10s and 1 s delay. Mean across the population of the absolute value of the difference between mean activity on information forced trials and mean activity on no information forced trials in each cell. Means calculated separately for sessions with 10s and 1s delay and on shuffled activity between the two conditions. 10s and 1s delay show s.e.m., shuffled data shows 1000 shuffled means. (h,i) Response to information and no information prediction at 10s and 1s delay. Mean activity of cells differentially responding to the (h) information and (i) no information forced trial center port odors calculated separately for sessions with 10s and 1s delay.

each comprising about 18% of the neurons, are distinct and exhibit only 18% overlap with each other (Fig4b). Moreover, little correlation is observed between the activity of these two representations (correlation = 0.15, p<0.0001, Fig4c). These data suggest that OFC maintains largely distinct CS+ representations that predict rewards of intrinsic and extrinsic value. Our behavioral experiments suggest that mice integrate these two populations to create a single value signal in driving choice decisions. The representation of this integrated value is therefore likely to result from convergence downstream of OFC.

We have described neural representations of a CS+ and US for the intrinsic value of information and a CS+ for the extrinsic value of water reward. We also observe a subset of OFC cells that responds to the receipt of water in either the Information or No Information port (FigS7, FigS6g,i). These cells may represent a US for the receipt of water. The US responses to the receipt of water and the complex response to odor A, the CS+ predictive of the receipt of water, were somewhat correlated (corr=0.12, p<0.001, Fig4d). However, we do not observe a correlation between the CS+ predictive of information and the receipt of water (Fig4e). Although they are intermixed across the full population of OFC cells, each of the subpopulations we identified displayed a particular pattern of activity across the trial, with differential encoding of information prediction, information receipt, water prediction, and water receipt (FigS6). These data provide further support for the existence of distinct representations of intrinsic information-related and extrinsic water-related value.

OFC signals reflect information value

Our behavioral experiments show that as we diminish the delay between the side odor predictive of reward and the receipt of water, mice diminish their preference for information. This suggests that the value of information concerning the receipt of reward is reduced with diminished delay, as observed in other species[16,18,29,56]. We therefore examined the OFC responses to odors at the center port that signal the receipt of information in reduced delay sessions (Fig5a). In accord with our behavioral data, in sessions in which the delay is reduced from 10s to 1s, we observed decreased responses of individual neurons to odors predictive of information. One example cell which displayed a decreased response to the information-predicting center port odor with decreased delay is shown in Fig.5b

Across the population, the number of cells comprising the information CS+ response was similar across both delays (Fig5c). However, these cells responded more weakly to both the information and no information-predicting center odors in sessions with a 1s delay, with the difference between responses to the center odors predicting information and no information diminished in sessions with 1s delay (Fig5d-l, FigS6). However, this did not reflect a general decrease in activity or differential signaling of trial events in 1s delay sessions (FigS8b-d). It also was not reflective of the passage of time between experiments conducted with 10s and 1s delays, as the information-no information difference was similar in sessions at with a comparable spacing in time but the same 10s delay (FigS8a). Rather, these data are in accord with our behavioral observation and modeling that mice show a decreased preference for the information port at shorter delays and suggest that the information CS+ representation reflects the subjective value of information as scaled by delay.

The emergence of a representation of information value with learning

We examined how OFC neural representations changed as an animal learned that information was available at one of the two side ports, an analysis enabled by our ability to register cellular identities across weeks of training[57]. During early training, the two different center port odors forced movement to the left or right port (there was no free choice option), where mice received equal water reward on every trial. Information about the receipt of reward was then introduced and water was delivered with 25% probability at both ports. Following this transition, animals learned that one port provided information as to whether they would receive water reward. We observed an increased differential response to the information-predicting odors after learning (Fig6a,b). Prior to the introduction of the informative odors, we observed many OFC cells that differentially responded to the left and right side-directing odors (Fig.6c). After information learning, many of these same cells displayed increased responses to the center port odor that directed the animal to the information side port. One such example cell is shown in Figure 6d. Across the population, we observe that the majority of cells that differentially respond to odors directing movement to the left or right side port responded differentially to the same odors after they became predictive of information or no information after learning (Fig 6e). After learning, the difference in the response to these two center port odors increased (Fig6a,b). This is due to an increased response across the population to the information-predicting odor after learning (Fig6c,f,g). Thus, a representation predictive of information only emerges upon learning. Moreover, while this representation of information emerges largely from the representation of odors directing the mouse to a particular side to receive water reward, it is robust to later reversals of side identity, when the mouse learns that information is now predicted by a different center port odor and available at a different side port location.

Low-dimensional model of information and water reward value

To further understand how this task and mouse behavior are represented in OFC neural activity, we used a nonlinear latent variable model called CEBRA (Consistent EmBeddings of high-dimensional Recordings using Auxiliary variables)[58]. In a complex task such as ours, activity from different or overlapping groups of neurons likely non-linearly combine to encode task/behavioral variables. Thus, how the task is encoded in the neural circuit at the level of aggregate neural dynamics, which can be studied in a low-dimensional

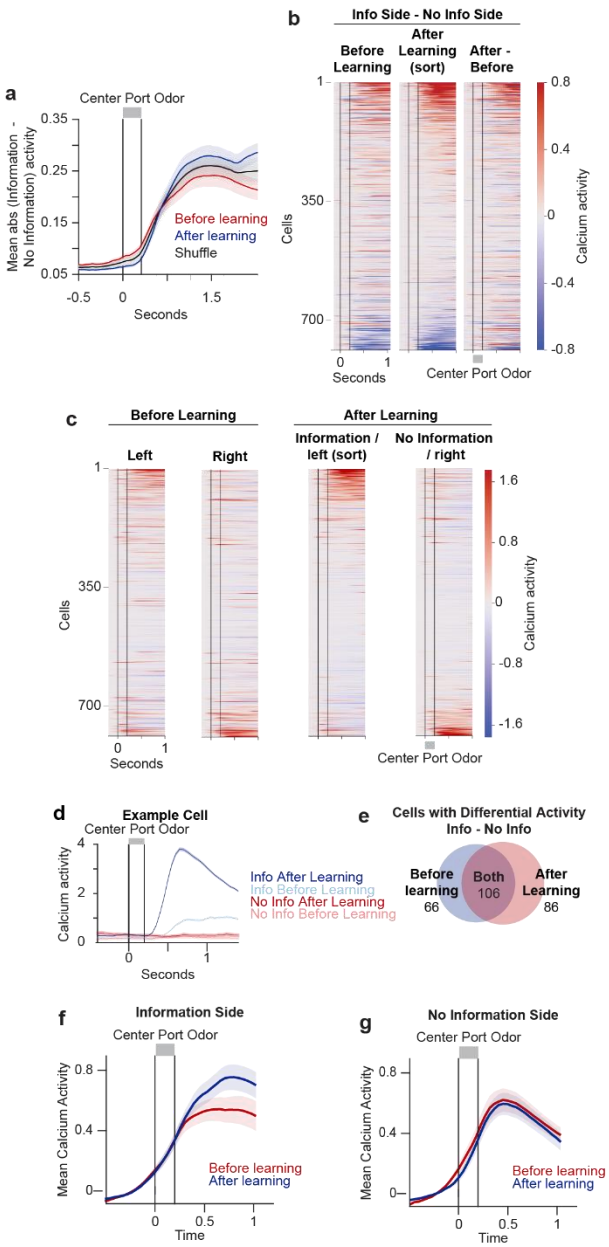

Figure 6: Emergence of a representation of value with learning

(a) Mean information CS+ representation before and after learning. Mean across the population of the absolute value of the difference between mean activity on information forced trials and mean activity on no information forced trials in each cell. Means calculated separately for sessions before and after the introduction of information to the task (learning), as indicated. Shuffled data shows 1000 shuffled means. (b) Population information CS+ representation before and after learning. Population activity of the differences in responses to the information and no information forced trial center port odor in sessions before and after learning. Right column shows the difference between the differences in responses to odors directing to the Information or No Information side port after learning – differences in responses to odors directing to the left or right side before learning. Each row shows the mean activity difference for each cell across all panels, sorted by the response difference after learning. (c) Responses to center port odors directing to the left or right side port before and after learning. Left, population activity of responses to the left or right forced trial center port odor in sessions before learning. Right, population activity of responses to the information (left) or no information (right) forced trial center port odor in sessions after learning. Each row shows the mean-subtracted activity for each cell across all panels, sorted by the information CS+ response, the difference in responses to information and no information. Note that left/right side port identity assignments are counterbalanced across animals, so the designation of "left" or "right" is arbitrary. (d) Responses of a single example cell to center port odors forcing movement to the left or right in sessions before and after learning of information. Information after learning is the same side as "left" before learning. (e) Number of cells with differential activity between the left and right and information and no information forced trial center port odors in sessions before and after learning, and their overlap. N=1138 total recorded cells. (f,g) Responses to left and right forced trial center odors before and after learning. Mean activity of cells active in response to the (f) left/information and (g) right/no information forced trial center port odors calculated separately for sessions before and after learning.

latent (embedding) space, can further enhance understanding of the information value encoding. Here, we applied CEBRA to our complex cognitive task to derive task-relevant low-dimensional embedding spaces by jointly fitting neural data and behavioral labels to accurately represent and decode task variables of interest from neural and behavioral data (see Methods).

CEBRA derives a common ("cohort-level") representation across all animals and recording sessions. In contrast, other (linear, unlabeled) dimensionality reduction methods[59–61] require aligning active neurons across different recording sessions prior to fitting and risk excluding potentially large numbers of recorded cells. Further, most alternative methods fit a separate model for data from each animal, making comparisons difficult and results harder to generalize. We used CEBRA to include all the recorded cells across recording sessions in the fitting process and derive a common embedding space across animals; while still making consistent inferences from data in a single animal, here, how complex variables such as *information* and *water value* are represented in OFC activity.

We fit a single cohort-level CEBRA model to neural recordings from six animals, with four imaging sessions each from our main dataset, for a total of 24 imaging sessions with approximately 400 recorded cells per session, on average. These sessions surrounded side reversals such that they included two sessions with information on the right, and two with information on the left, for each animal. We fit the neural data jointly with behavioral labels for information choice, water reward, odor presentation times, reaction times and trial time (see Methods). Importantly, the model did not include session-level labels for left and right choice, and thus, learned a representation that was independent of side. The model was fit to the activity of all recorded cells within a session without registering cell identities across days and, therefore, our CEBRA-based analyses provide further evidence for the existence of these distinct

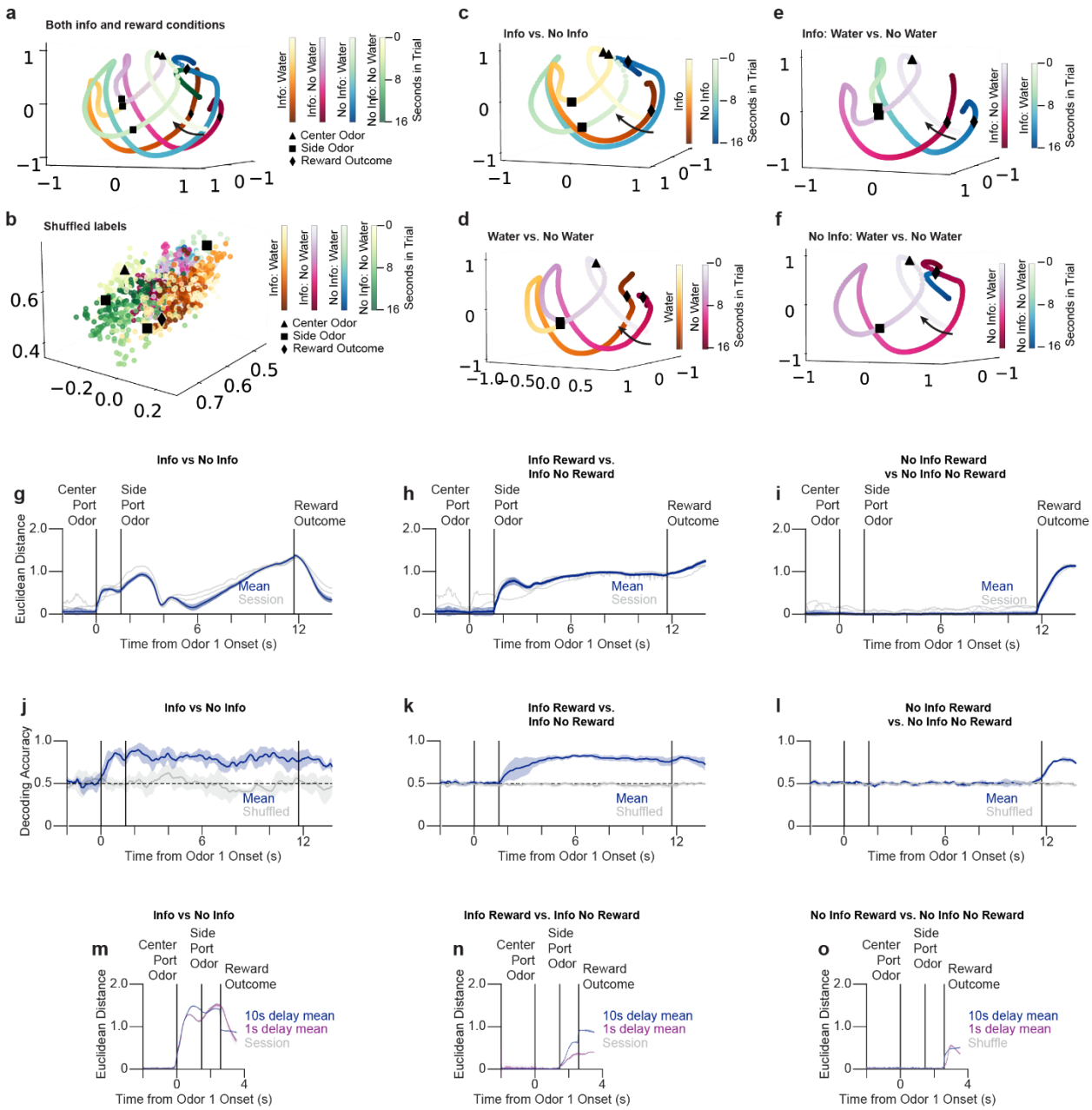

Figure 7: CEBRA low-dimensional model of information and water reward value

(a) Plot of the first three latents of CEBRA embedding across trial time for interacting labels of information and reward. (b) The same plot for data with maximally shuffled labels (see Methods for details). (c-f) Plots of the first three latents of CEBRA embeddings for (c) information versus no information and (d) reward versus unrewarded trial trajectories in state space, then interaction term trajectories of (e) rewarded versus unrewarded information trials and (f) rewarded versus unrewarded no information trials. Quantitative analysis of the embedding latents' coordinates was then performed using Euclidean distance calculations between each type of behavior-labeled trajectory at each frame over trial-time: (g) the distance between information and no information trajectories, and between rewarded and unrewarded trajectories for (h) information trials and (i) no information trials. (j-l) To test the quality of these embeddings, as well as to find at which trial periods representations of behavioral variables were encoded in the OFC, decoding accuracy was then calculated and compared between decoding trial variables (information versus no information, and reward-information interactions), from CEBRA manifold embeddings. All plots up to this point were for embeddings fit only to 10 s delay trials. Last, (m-o) these same distances in (g-i) were also calculated in a joint model that fit both 1 s delay trials and a truncated version of 10 s delay trials with ~9 sec of the delay removed to align the first 1 s of the delay and reward periods between each delay category. Error bars in CEBRA decoding and distance plots are standard error.

representations independent of the fidelity of cell registration. We find highly consistent structure across the cohort for context-labeled CEBRA projections of data from individual sessions and animals, with low variation between single mice context-labeled trajectory differences (gray lines in Fig7g-i). This structure collapses in label-shuffled controls (Fig7a-f, FigS9).

We examined whether representations of intrinsic information and extrinsic water value were dissociable in the CEBRA-derived task embedding. To that end, we projected OFC activity across trials split by information and water reward in the latent embedding space, and examined trial-averaged dynamics across time (Fig 7a-f). The trajectories of different task variables in the embedding space reflect the representation (or encoding) of these variables in the aggregate dynamics of the OFC neural population throughout the course of the trial. The representations of information and no information trials diverged immediately after the center odor presentation (Fig 7c). The representations of reward and no reward on information trials diverged immediately after side odor presentations and separated further following the reward outcome (Fig7e); on no information trials, representations diverged only after the reward outcome period (Fig 7f).

We quantified the separation between the latent embeddings of these task variables at each moment in time by computing the Euclidean distance between them in the embedding space. The distance between the trajectories of information versus no information trial activity increased sharply following the center odor presentation and continued to increase through the presentation of the side odors (Fig7g), consistent with OFC containing a representation of both the prediction and receipt of information. Further, the distance between trajectories split by rewarded versus unrewarded trials increased following the side odor presentation on information trials and the reward outcome on no information trials (Fig7h,i). The divergence of the task representations at the appropriate times suggests that the model faithfully encodes the task structure and that these task variables are encoded by OFC dynamics at the relevant times.

We therefore asked whether the differences between information and no information and reward and no reward could be decoded from the low-dimensional CEBRA manifold embeddings. Decoding of information versus no information from the CEBRA manifold was highly accurate following the center odor presentation and throughout the rest of the trial (Fig7j). Reward versus no reward was decodable following the presentation of the side odor and increased at the reward outcome, consistent with reward being revealed at the side odor on the information side and not until outcome on the no information side (Fig7k,l). Additional shuffled controls were performed to support the validity of the CEBRA model and decoder, finding expected decoder performance deterioration with increased label shuffling and reduced sample size (FigS9). The ability to decode information versus no information, and reward versus no reward trials, from the CEBRA model fit across all animals suggests that these two kinds of value are generally represented in aggregate OFC neural dynamics and are decodable from a low-dimensional manifold during performance of the information seeking task.

We next modeled OFC activity across sessions with 1s versus 10s delays between information and outcome. We truncated the delay period following the side odor presentation in the 10s trials by removing the later 9s of the delay and aligned the timing of the reward outcome periods to allow a model to be jointly fit across these sessions of different trial length. Confirming our single-cell-based analysis, the distance between information and no information trajectories was diminished in sessions with a 1s delay compared to those with a 10s delay (Fig7m). Thus, the information and no-information representations were more separated in state space on the 10s delay trials than in the 1s delay trials, suggesting that the value of information encoded by OFC is indeed scaled by the delay length between cue and reward as proposed in earlier work across species[6,16,19,49]. Further, the separation between rewarded and unrewarded trials that emerges after the side odor on the information side was larger with longer delay, a finding we also observed at the single cell level (Fig7n,FigS8). Thus, the representational difference between information-providing odors A and B is greater with greater information value and longer delay.

DISCUSSION

We have developed a behavioral paradigm to demonstrate that mice choose to acquire information even though this information does not alter the reward outcome of the task. Moreover, mice are willing to sacrifice reward for this knowledge suggesting that in mice, as in other animals, knowledge is of intrinsic value[5,7,9,13,14,16–20,28]. Calcium imaging during the information task reveals different but overlapping populations of neurons responsive to the odor predictive of information and odors predictive of water reward. In addition, we observed distinct representations of neurons responsive to odors that signal the receipt of information and water reward itself. These data suggest that mice have evolved distinct pathways in OFC that represent the intrinsic value of knowledge and the extrinsic value of water reward. The presence of distinct representations within OFC may allow the organism to separately learn and update different elements of a rewarding stimulus and execute different behaviors in response to distinct value representations. Our behavioral data, however, reveal that mice are willing to sacrifice water for information indicating that the subjective value of information and the objective value of water are ultimately integrated to elicit an appropriate behavioral response.

Electrophysiological recordings in monkeys during an information seeking task conceptually analogous to our paradigm similarly reveal orthogonal coding of information and water reward in OFC[23]. Recordings in the monkey have also identified a cortex-basal ganglia circuit responsive to information and extrinsic reward[4,29]. Anterior cingulate cortex (ACC) and downstream areas in striatum and pallidum signal uncertainty and respond to the receipt of information by diminished firing,

presumably reflecting the resolution of uncertainty. The information-responsive circuits and extrinsic reward circuits in the cortex-basal ganglia loops are distinct and do not appear to fully integrate intrinsic and extrinsic reward[7,29]. However, integration is observed downstream of pallidum, in the lateral habenula and the dopamine neurons of the ventral tegmental area[28,29]. The activity of these neurons more accurately reflect subjective value and integrate the value of information with extrinsic reward.

Our paradigm has enabled the study of information seeking behavior in mice, allowing the use of methods that are not available in other information seeking species such as monkeys or humans. Miniscope imaging during our behavioral paradigm allowed us to follow the emergence during task training of representations that encode the value of cues predictive of information as well as odor cues that provide information at the side port. One advantage of our task compared to other studies of information seeking is our use of odor cues to both predict and provide information. The emergence of a representation predictive of information in OFC in response to the center port odors implies that the odor input from piriform cortex to OFC is modified during task learning[47]. Mice must recognize that the odors in one side port provide information in order to learn that the center port odor predicts information. We observed highly overlapping representations of the two side odors, A and B, at the information port. Both odors signal the receipt of information, one signaling the presence of water reward and the other the absence of reward. This suggests that these odor cues serve as unconditioned stimuli signaling the receipt of information and may reinforce the representation of odor cues predictive of information at the center port.

This is a relatively complex process of cognitive abstraction because recognition that an odor provides information requires knowledge of its predictive nature despite the lack of its association with an external reward. This observation poses the question as to how the brain initially recognizes that an odor cue provides information of subjective value. Models of classical conditioning involving rewards of extrinsic value, such as food or water, invoke known physiologic responses that activate subcortical brain structures that lead to the release of dopamine and downstream reinforcement[62]. Analogous models for rewards of intrinsic value suggest that in our information seeking task there must be an odor-responsive neural ensemble that encodes the value of information and is capable of reinforcing information-predictive odor cues. However, the nature of the brain structures capable of recognizing the innate value of knowledge and activating both the cortex-basal ganglia loop and distinct OFC subpopulations remain elusive. In the OFC we observe distinct populations of cells, one representing the predicted value of information and a second representing the predicted value of water. Although dopamine may reinforce both neural populations, one population can only be reinforced in the context of information and the second only in the context of water.

The desire for knowledge of no apparent extrinsic value was implicit in the early behavioral experiments of Tolman[1]. These studies revealed that rats develop, without obvious reinforcement, a cognitive spatial map which is revealed when they appear to use their prior knowledge to navigate the maze when it yields reward. The Tolman experiment is instructive because it not only reveals innate acquisition of information about the world but also suggests how this seemingly innate drive may have evolved. At the outset of the experiment there is no apparent extrinsic value to the learning of a cognitive spatial map. When food is later placed in the maze, however, this prior knowledge becomes valuable. The recognition that the acquisition of knowledge of no obvious immediate extrinsic value may ultimately enhance later extrinsic value affords a selective advantage to the “desire to know”. The drive to acquire information is observed in experiments with animals from pigeons to humans suggesting the conservation of innate neural pathways that impose intrinsic value on this knowledge.

Why does an organism value information of no apparent extrinsic value? One class of models argues that information seeking is a consequence of secondary reinforcement[19,63–67]. The receipt of odor A (the odor predictive of water reward) at the information side port may be overvalued and secondarily reinforce the choice of information at the center port. A second model posits that odor A at the side port boosts “anticipatory delight” enhancing the value of the information port[16]. These models may not easily account for observations that humans and other animals exhibit a desire for information about upcoming aversive events and events that have already occurred[5,14,68–70]. Others have argued that the resolution of uncertainty provides a basis for the value of information[29,69,71–75]. In many contexts organisms experience uncertainty as an aversive state perhaps analogous to the aversion experienced during hunger or thirst[19,76,77]. Unlike physiologic states elicited by homeostatic body chemistry, uncertainty remains an elusive cognitive state. Perhaps the simplest explanation argues that information led, at some time during the evolution of a species, to extrinsic value. Indeed, in natural environments, knowledge improves an organism’s model of the world, and its acquisition improves decision making and augments survival[78]. This quest for information may therefore have been fixed by selection and now generalizes to circumstances in which the acquisition of knowledge may be of no extrinsic value. This model suggests that the desire for knowledge evolved by assigning innate pleasure to the acquisition of information. Thus, the value of information may ultimately reflect the innate pleasure it provides. As Dante posits in his Convivio, “Since knowledge is the ultimate perfection of our soul, in which resides our ultimate happiness, we are all therefore by nature subject to a desire for it”.

AUTHOR CONTRIBUTIONS

J.J.B. developed the information seeking task, trained animals and collected behavior data, performed imaging

experiments, and performed single-cell neural analyses. R.P.B., C.D.M., J.J.B., and K.R. conceptualized and implemented the behavioral decision and the CEBRA model. E.S.B.-M. developed the information seeking task. L.F.A. supervised single-cell neural analyses. K.R. supervised behavioral and population neural modeling and CEBRA analysis. R.A. supervised all stages of the project. J.J.B. and R.A. wrote the manuscript. All authors contributed to the editing of the manuscript and approved the final version.

ACKNOWLEDGEMENTS

The authors would like to thank Evan Schaffer for his suggestions on cell registration and motion correction. They would like to thank members of the Axel lab, Ilya Monosov, Mackenzie Mathis, and Jackie Gottlieb for helpful discussions. They would also like to thank the research assistants Maya Sharma Campbell, Ashwin Viswanathan, Deniz Bingul, Theodore Hannah, and Adamu Awak who performed mouse behavioral training as well as Nataliya Zabello and the staff of the Zuckerman Institute for Comparative Medicine for their expert mouse care. We would also like to acknowledge Tanya Tabachnik and the Zuckerman Institute Advanced Instrumentation Core for their help with design of behavioral chambers. This work is supported by NIH RF1 DA056403 (R.P.B., C.D.M., K.R.), an NSF CAREER Award (K.R.), the James S. McDonnell Foundation (K.R.), the McKnight Foundation (K.R.), the CIFAR (R.P.B., C.D.M., K.R.), the Simons Foundation (J.J.B., L.F.A., R.P.B., C.D.M., K.R.), the Gatsby Charitable Foundation (L.F.A.), and the Howard Hughes Medical Institute (R.A.). R.A. is an HHMI investigator.

REFERENCES

1. Tolman, E.C., and Honzik, C.H. (1930). Introduction and removal of reward, and maze performance in rats. Univ. Calif. Publ. Psychol. *4*, 257–275.
2. Prokasy Jr., W.F. (1956). The acquisition of observing responses in the absence of differential external reinforcement. J. Comp. Physiol. Psychol. *49*, 131–134. 10.1037/h0046740.
3. Charpentier, C.J., and Dezza, I.C. (2021). Information-seeking in the brain. 10.31234/osf.io/qfxgd.
4. White, J.K., Bromberg-Martin, E.S., Heilbronner, S.R., Zhang, K., Pai, J., Haber, S.N., and Monosov, I.E. (2019). A neural network for information seeking. Nat. Commun. *10*, 1–19. 10.1038/s41467-019-13135-z.
5. Kobayashi, K., and Hsu, M. (2019). Common neural code for reward and information value. Proc. Natl. Acad. Sci. *116*, 13061–13066. 10.1073/pnas.1820145116.
6. Iigaya, K., Hauser, T.U., Kurth-Nelson, Z., O'Doherty, J.P., Dayan, P., and Dolan, R.J. (2020). The value of what's to come: Neural mechanisms coupling prediction error and the utility of anticipation. Sci. Adv. *6*, eaba3828. 10.1126/sciadv.aba3828.
7. Dezza, I.C., Cleeremens, A., and Alexander, W. (2021). Independent and Interacting Value Systems for Reward and Information in the Human Brain 10.1101/2020.05.04.075739.
8. Horan, M., Daddaoua, N., and Gottlieb, J. (2019). Parietal neurons encode information sampling based on decision uncertainty. Nat. Neurosci., 1. 10.1038/s41593-019-0440-1.
9. González, V.V., Ashikyan, S.A., Zhang, Y., Rickard, A., Yassine, I., Romero-Sosa, J.L., Blaisdell, A.P., and Izquierdo, A. (2023). A special role for Anterior Cingulate Cortex, but not Orbitofrontal Cortex or Basolateral Amygdala, in choices involving information. 2023.08.03.551514. 10.1101/2023.08.03.551514.
10. Filimon, F., Nelson, J.D., Sejnowski, T.J., Sereno, M.I., and Cottrell, G.W. (2020). The ventral striatum dissociates information expectation, reward anticipation, and reward receipt. Proc. Natl. Acad. Sci. *117*, 15200–15208. 10.1073/pnas.1911778117.
11. Eschmann, K.C.J., Pereira, D.F.M.M., Valji, A., Dehmelt, V., and Gruber, M.J. (2022). Curiosity and mesolimbic functional connectivity drive information seeking in real life. 2022.01.28.478038. 10.1101/2022.01.28.478038.
12. Foley, N.C., Kelly, S.P., Mhatre, H., Lopes, M., and Gottlieb, J. (2017). Parietal neurons encode expected gains in instrumental information. Proc. Natl. Acad. Sci., 201613844. 10.1073/pnas.1613844114.
13. Li, Y., Daddaoua, N., Horan, M., Foley, N.C., and Gottlieb, J. (2022). Uncertainty modulates visual maps during noninstrumental information demand. Nat. Commun. *13*, 5911. 10.1038/s41467-022-33585-2.
14. Charpentier, C.J., Bromberg-Martin, E.S., and Sharot, T. (2018). Valuation of knowledge and ignorance in mesolimbic reward circuitry. Proc. Natl. Acad. Sci., 201800547. 10.1073/pnas.1800547115.
15. Jezzini, A., Bromberg-Martin, E.S., Trambaiolli, L.R., Haber, S.N., and Monosov, I.E. (2021). A prefrontal network integrates preferences for advance information about uncertain rewards and punishments. Neuron, S0896627321003536. 10.1016/j.neuron.2021.05.013.
16. Iigaya, K., Story, G.W., Kurth-Nelson, Z., Dolan, R.J., and Dayan, P. (2016). The modulation of savouring by prediction error and its effects on choice. eLife *5*, e13747. 10.7554/eLife.13747.
17. Roper, K.L., and Zentall, T.R. (1999). Observing behavior in pigeons: The effect of reinforcement probability and response cost using a symmetrical choice procedure. Learn. Motiv. *30*, 201–220. 10.1006/lmot.1999.1030.
18. McDevitt, M.A., Pisklak, J.M., and Spetch, M. (2018). The Influence of Outcome Delay on Suboptimal Choice. Behav. Processes. 10.1016/j.beproc.2018.10.008.
19. Ajuwon, V., Ojeda, A., Murphy, R.A., Monteiro, T., and Kacelnik, A. (2022). Paradoxical choice and the reinforcing value of information. Anim. Cogn. 10.1007/s10071-022-01698-2.

20. Ojeda, A., Murphy, R.A., and Kacelnik, A. (2018). Paradoxical choice in rats: Subjective valuation and mechanism of choice. Behav. Processes. 10.1016/j.beproc.2018.03.024.
21. Zhu, J.-Q., Xiang, W., and Ludvig, E.A. (2017). Information seeking as chasing anticipated prediction errors. In Proceedings of the 39th Annual Meeting of the Cognitive Science Society.
22. Roper, K.L., and Baldwin, E.R. (2004). The two-alternative observing response procedure in rats: Preference for nondiscriminative stimuli and the effect of delay. Learn. Motiv. *35*, 275–302. 10.1016/j.lmot.2004.04.004.
23. Blanchard, T.C., Hayden, B.Y., and Bromberg-Martin, E.S. (2015). Orbitofrontal Cortex Uses Distinct Codes for Different Choice Attributes in Decisions Motivated by Curiosity. Neuron. 10.1016/j.neuron.2014.12.050.
24. Bennett, D., Bode, S., Brydevall, M., Warren, H., and Murawski, C. (2016). Intrinsic Valuation of Information in Decision Making under Uncertainty. PLOS Comput. Biol. *12*, e1005020. 10.1371/journal.pcbi.1005020.
25. Rodriguez Cabrero, J.A.M., Zhu, J.-Q., and Ludvig, E.A. (2019). Costly curiosity: People pay a price to resolve an uncertain gamble early. Behav. Processes *160*, 20–25. 10.1016/j.beproc.2018.12.015.
26. Cunningham, P.J., and Shahan, T.A. (2019). Rats engage in suboptimal choice when the delay to food is sufficiently long. J. Exp. Psychol. Anim. Learn. Cogn. *45*, 301–310. 10.1037/xan0000211.
27. Murayama, K. (2022). A reward-learning framework of knowledge acquisition: An integrated account of curiosity, interest, and intrinsic–extrinsic rewards. Psychol. Rev., No Pagination Specified-No Pagination Specified. 10.1037/rev0000349.
28. Bromberg-Martin, E.S., and Hikosaka, O. (2009). Midbrain Dopamine Neurons Signal Preference for Advance Information about Upcoming Rewards. Neuron *63*, 119–126. 10.1016/j.neuron.2009.06.009.
29. Bromberg-Martin, E.S., Feng, Y.-Y., Ogasawara, T., White, J.K., Zhang, K., and Monosov, I.E. (2022). A neural mechanism for conserved value computations integrating information and rewards. 2022.08.14.503903. 10.1101/2022.08.14.503903.
30. Hayden, B., and Niv, Y. (2020). The case against economic values in the brain (PsyArXiv) 10.31234/osf.io/7hgup.
31. Padoa-Schioppa, C., and Assad, J.A. (2006). Neurons in the orbitofrontal cortex encode economic value. Nature *441*, 223–226. 10.1038/nature04676.
32. Ballesta, S., Shi, W., Conen, K.E., and Padoa-Schioppa, C. (2020). Values encoded in orbitofrontal cortex are causally related to economic choices. Nature *588*, 450–453. 10.1038/s41586-020-2880-x.
33. Kuwabara, M., Kang, N., Holy, T.E., and Padoa-Schioppa, C. (2020). Neural mechanisms of economic choices in mice. eLife *9*, e49669. 10.7554/eLife.49669.
34. Wilson, R.C., Takahashi, Y.K., Schoenbaum, G., and Niv, Y. (2014). Orbitofrontal Cortex as a Cognitive Map of Task Space. Neuron *81*, 267–279. 10.1016/j.neuron.2013.11.005.
35. Stalnaker, T.A., Cooch, N.K., and Schoenbaum, G. (2015). What the orbitofrontal cortex does not do. Nat. Neurosci. *18*, 620–627. 10.1038/nn.3982.
36. Kable, J.W., and Glimcher, P.W. (2007). The neural correlates of subjective value during intertemporal choice. Nat. Neurosci. *10*, 1625–1633. 10.1038/nn2007.
37. Ottenheimer, D.J., Hjort, M.M., Bowen, A.J., Steinmetz, N.A., and Stuber, G.D. (2023). A stable, distributed code for cue value in mouse cortex during reward learning. eLife *12*, RP84604. 10.7554/eLife.84604.
38. Cai, X., Kim, S., and Lee, D. (2011). Heterogeneous Coding of Temporally Discounted Values in the Dorsal and Ventral Striatum during Intertemporal Choice. Neuron *69*, 170–182. 10.1016/j.neuron.2010.11.041.
39. Constantinople, C.M., Piet, A.T., Bibawi, P., Akrami, A., Kopec, C., and Brody, C.D. (2019). Lateral orbitofrontal cortex promotes trial-by-trial learning of risky, but not spatial, biases. eLife *8*, e49744. 10.7554/eLife.49744.
40. Gore, F., Hernandez, M., Ramakrishnan, C., Crow, A.K., Malenka, R.C., and Deisseroth, K. (2023). Orbitofrontal cortex control of striatum leads economic decision-making. Nat. Neurosci., 1–9. 10.1038/s41593-023-01409-1.
41. Feierstein, C.E., Quirk, M.C., Uchida, N., Sosulski, D.L., and Mainen, Z.F. (2006). Representation of Spatial Goals in Rat Orbitofrontal Cortex. Neuron *51*, 495–507. 10.1016/j.neuron.2006.06.032.
42. Costa, K.M., Scholz, R., Lloyd, K., Moreno-Castilla, P., Gardner, M., Dayan, P., and Schoenbaum, G. (2022). The role of the orbitofrontal cortex in creating cognitive map. 10.1101/2022.01.25.477716.
43. Gardner, M.P.H., Sanchez, D., Conroy, J.C., Wikenheiser, A.M., Zhou, J., and Schoenbaum, G. (2020). Processing in Lateral Orbitofrontal Cortex Is Required to Estimate Subjective Preference during Initial, but Not Established, Economic Choice. Neuron *108*, 526-537.e4. 10.1016/j.neuron.2020.08.010.
44. Miller, K.J., Botvinick, M.M., and Brody, C.D. (2018). Value Representations in Orbitofrontal Cortex Drive Learning, but not Choice. bioRxiv, 245720. 10.1101/245720.
45. van Lieshout, L.L.F., Vandenbroucke, A.R.E., Müller, N.C.J., Cools, R., and de Lange, F.P. (2018). Induction and Relief of Curiosity Elicit Parietal and Frontal Activity. J. Neurosci. *38*, 2579–2588. 10.1523/JNEUROSCI.2816-17.2018.
46. Kaanders, P., Nili, H., O'Reilly, J.X., and Hunt, L. (2021). Medial Frontal Cortex Activity Predicts Information Sampling in Economic Choice. J. Neurosci. *41*, 8403–8413. 10.1523/JNEUROSCI.0392-21.2021.
47. Wang, P.Y., Boboila, C., Chin, M., Higashi-Howard, A., Shamash, P., Wu, Z., Stein, N.P., Abbott, L.F., and Axel, R. (2020). Transient and Persistent Representations of Odor Value in Prefrontal Cortex.

Neuron *108*, 209-224.e6. 10.1016/j.neuron.2020.07.033.
48. Eliaz, K., and Schotter, A. (2010). Paying for confidence: An experimental study of the demand for non-instrumental information. Games Econ. Behav. *70*, 304–324. 10.1016/j.geb.2010.01.006.
49. Bromberg-Martin, E.S., and Monosov, I.E. (2020). Neural circuitry of information seeking. Curr. Opin. Behav. Sci. *35*, 62–70. 10.1016/j.cobeha.2020.07.006.
50. Rescorla, R.A., and Wagner, A.R. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Classical Conditioning II: Current Research and Theory (Appleton-Century-Crofts), pp. 64–99.
51. Dezza, I.C., Yu, A.J., Cleeremans, A., and Alexander, W. (2017). Learning the value of information and reward over time when solving exploration-exploitation problems. Sci. Rep. *7*, 16919. 10.1038/s41598-017-17237-w.
52. Padoa-Schioppa, C. (2007). Orbitofrontal Cortex and the Computation of Economic Value. Ann. N. Y. Acad. Sci. *1121*, 232–253. 10.1196/annals.1401.011.
53. Schoenbaum, G., Takahashi, Y., Liu, T.-L., and McDannald, M.A. (2011). Does the orbitofrontal cortex signal value? Ann. N. Y. Acad. Sci. *1239*, 87–99. 10.1111/j.1749-6632.2011.06210.x.
54. Ramus, S.J., and Eichenbaum, H. (2000). Neural Correlates of Olfactory Recognition Memory in the Rat Orbitofrontal Cortex. J. Neurosci. *20*, 8199–8208. 10.1523/JNEUROSCI.20-21-08199.2000.
55. Fusi, S., Miller, E.K., and Rigotti, M. (2016). Why neurons mix: high dimensionality for higher cognition. Curr. Opin. Neurobiol. *37*, 66–74. 10.1016/j.conb.2016.01.010.
56. Vasconcelos, M., Monteiro, T., and Kacelnik, A. (2015). Irrational choice and the value of information. Sci. Rep. *5*, 13874.
57. Giovannucci, A., Friedrich, J., Gunn, P., Kalfon, J., Brown, B.L., Koay, S.A., Taxidis, J., Najafi, F., Gauthier, J.L., Zhou, P., et al. (2019). CaImAn an open source tool for scalable calcium imaging data analysis. eLife *8*, e38173. 10.7554/eLife.38173.
58. Schneider, S., Lee, J.H., and Mathis, M.W. (2023). Learnable latent embeddings for joint behavioural and neural analysis. Nature *617*, 360–368. 10.1038/s41586-023-06031-6.
59. Kobak, D., Brendel, W., Constantinidis, C., Feierstein, C.E., Kepecs, A., Mainen, Z.F., Qi, X.-L., Romo, R., Uchida, N., and Machens, C.K. (2016). Demixed principal component analysis of neural population data. eLife *5*, e10989. 10.7554/eLife.10989.
60. Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. J. Educ. Psychol. *24*, 417–441. 10.1037/h0071325.
61. Pang, R., Lansdell, B.J., and Fairhall, A.L. (2016). Dimensionality reduction in neuroscience. Curr. Biol. *26*, R656–R660. 10.1016/j.cub.2016.05.029.
62. Grove, J.C.R., Gray, L.A., La Santa Medina, N., Sivakumar, N., Ahn, J.S., Corpuz, T.V., Berke, J.D., Kreitzer, A.C., and Knight, Z.A. (2022). Dopamine subsystems that track internal states. Nature *608*, 374–380. 10.1038/s41586-022-04954-0.
63. Lieberman, D.A. (1972). Secondary reinforcement and information as determinants of observing behavior in monkeys (Macaca mulatta). Learn. Motiv. *3*, 341–358. 10.1016/0023-9690(72)90030-6.
64. Beierholm, U.R., and Dayan, P. (2010). Pavlovian-Instrumental Interaction in 'Observing Behavior.' PLOS Comput. Biol. *6*, e1000903. 10.1371/journal.pcbi.1000903.
65. Dunn, R.M., Pisklak, J.M., McDevitt, M.A., and Spetch, M.L. (2023). Suboptimal choice: A review and quantification of the signal for good news (SiGN) model. Psychol. Rev. 10.1037/rev0000416.
66. Chow, J.J., Smith, A.P., Wilson, A.G., Zentall, T.R., and Beckmann, J.S. (2017). Suboptimal choice in rats: Incentive salience attribution promotes maladaptive decision-making. Behav. Brain Res. *320*, 244–254. 10.1016/j.bbr.2016.12.013.
67. Fantino, E., and Moore, J. (1980). Uncertainty Reduction, Conditioned Reinforcement, and Observing. J. Exp. Anal. Behav. *33*, 3–13. 10.1901/jeab.1980.33-3.
68. FitzGibbon, L., Komiya, A., and Murayama, K. (2019). The Lure of Counterfactual Curiosity: People Incur a Cost to Experience Regret (Open Science Framework) 10.31219/osf.io/jm3uc.
69. Wang, M.Z., and Hayden, B.Y. (2019). Monkeys are curious about counterfactual outcomes. Cognition *189*, 1–10. 10.1016/j.cognition.2019.03.009.
70. Marvin, C.B., and Shohamy, D. (2016). Curiosity and reward: Valence predicts choice and information prediction errors enhance learning. J. Exp. Psychol. Gen. *145*, 266–272. 10.1037/xge0000140.
71. Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., and Pezzulo, G. (2015). Active inference and epistemic value. Cogn. Neurosci. *6*, 187–214. 10.1080/17588928.2015.1020053.
72. Berlyne, D.E. (1962). Uncertainty and Epistemic Curiosity. Br. J. Psychol. *53*, 27–34. 10.1111/j.2044-8295.1962.tb00811.x.
73. Berlyne, D.E. (1950). Novelty and Curiosity as Determinants of Exploratory Behaviour. Br. J. Psychol. Gen. Sect. *41*, 68–80. 10.1111/j.2044-8295.1950.tb00262.x.
74. Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. Psychol. Bull. *116*, 75–98. 10.1037/0033-2909.116.1.75.
75. Golman, R., and Loewenstein, G. (2015). Curiosity, Information Gaps, and the Utility of Knowledge (Social Science Research Network).
76. Monosov, I.E. (2020). How Outcome Uncertainty Mediates Attention, Learning, and Decision-Making. Trends Neurosci. *43*, 795–809. 10.1016/j.tins.2020.06.009.
77. Bar-Anan, Y., Wilson, T.D., and Gilbert, D.T. (2009). The feeling of uncertainty intensifies affective reactions. Emotion *9*, 123–127. 10.1037/a0014607.

78. Behrens, T.E.J., Woolrich, M.W., Walton, M.E., and Rushworth, M.F.S. (2007). Learning the value of information in an uncertain world. Nat. Neurosci. *10*, 1214–1221. 10.1038/nn1954.
79. Daw, N.D. (2011). Trial-by-trial data analysis using computational models: (Tutorial Review). In Decision Making, Affect, and Learning: Attention and Performance XXIII, M. R. Delgado, E. A. Phelps, and T. W. Robbins, eds. (Oxford University Press), p. 0. 10.1093/acprof:oso/9780199600434.003.0001.
80. Eckstein, M.K., Wilbrecht, L., and Collins, A.G. (2021). What do reinforcement learning models measure? Interpreting model parameters in cognition and neuroscience. Curr. Opin. Behav. Sci. *41*, 128–137. 10.1016/j.cobeha.2021.06.004.
81. Wilson, R.C., and Collins, A.G. (2019). Ten simple rules for the computational modeling of behavioral data. eLife *8*, e49547. 10.7554/eLife.49547.
82. Akaishi, R., Umeda, K., Nagase, A., and Sakai, K. (2014). Autonomous Mechanism of Internal Choice Estimate Underlies Decision Inertia. Neuron *81*, 195–206. 10.1016/j.neuron.2013.10.018.
83. Akaike, H. (1974). A new look at the statistical model identification. IEEE Trans. Autom. Control *19*, 716–723. 10.1109/TAC.1974.1100705.
84. Pnevmatikakis, E.A., and Giovannucci, A. (2017). NoRMCorre: An online algorithm for piecewise rigid motion correction of calcium imaging data. J. Neurosci. Methods *291*, 83–94. 10.1016/j.jneumeth.2017.07.031.
85. Zhou, P., Resendez, S.L., Rodriguez-Romaguera, J., Jimenez, J.C., Neufeld, S.Q., Giovannucci, A., Friedrich, J., Pnevmatikakis, E.A., Stuber, G.D., Hen, R., et al. (2018). Efficient and accurate extraction of in vivo calcium signals from microendoscopic video data. eLife. 10.7554/eLife.28728.
86. Oord, A. van den, Li, Y., and Vinyals, O. (2019). Representation Learning with Contrastive Predictive Coding. 10.48550/arXiv.1807.03748.

.

METHODS

Subjects

All experimental and surgical protocols were performed in accordance with the guide of Care and Use of Laboratory Animals (NIH) and were approved by the Institutional Animal Care and Use Committee at Columbia University. Mice were water restricted at least 2 days prior to the initiation of training and maintained at >80% of their initial body weight throughout experiments. Mice were wild type C57/BL6 (Jackson Laboratories strain 000664), aged 8-45 weeks, of both male and female sex. All animals were maintained under reverse cycled 12hr light/12hr dark conditions and single housed after beginning water restriction. Experiments were performed during the dark period of the circadian cycle.

Information seeking task

In general, mice were trained on the information seeking task for 45-90min each weekday. Training sessions lasted until mice received at least 700uL of water or stopped engaging in trials. On average, mice received 800-1400uL of water per session and received water later that day in their home cage if daily requirements were not met during training.

Main task

Mice were trained to perform a 2-alternative-forced-choice task in which they chose whether to receive information revealing the trial's reward outcome (water or no water). Behavioral training and testing were performed in custom-made acrylic boxes outfitted with three nosepoke ports (Sanworks) along one wall in custom sound attenuating enclosures. In the information seeking task, mice initiated trials by poking in a center nosepoke port where one of three trial type odor cues was provided for 200ms. Following a 50ms 4500Hz go cue tone, mice then chose either Information or no information by nosepoking either the left or right reward port. Each animal was initially assigned either left or right as the Information port. Mice had to indicate their choice prior to the delivery of a second 200ms odor cue in the chosen side port 1.2s after the completion of the go cue or the trial had to be repeated after allowing the remainder of its duration to elapse. On some training sessions a grace period of up to 10s was provided to extend the time available for the mouse to indicate its choice. In this case, the side odor was presented as soon as the mouse entered the correct side port. If mice did not stay in the center nosepoke for the duration of the center odor cue and go cue, they were unable to proceed with the trial and had to re-poke the center port for the full odor duration to proceed. Trials were of three types: Forced Information, Forced No Information, and Choice. If mice chose the incorrect port on forced trials, they did not receive side port odor or reward and experienced a timeout for the remainder of the full trial time and had to repeat the trial. These procedures ensured mice were unable to avoid non-preferred trial types and equalized reward rates across trial types during training. At the side ports, on each trial mice received one of four odors for 200ms: the Information port provided odor A on 25% of trials, which was always followed by water reward, or odor B on 75% of trials, which was never followed by water reward. The No Information port provided odor C on 75% of trials and odor D on 25% of trials, but the water reward was determined independently and provided on 25% of No Information trials. The delivery of the side odor was followed by a delay period of 10s in the main task. A 4000Hz 200ms tone then indicated the outcome time on all trials, although water was only provided at that time on rewarded trials. Water rewards

in the main task were 16uL provided as 4-4uL drops 50ms apart. The reward outcome period was followed by a 3s inter-trial interval. Mice were required to be present in the chosen port for water to be delivered, but they were not otherwise required to be in the port after indicating their choice.

All odors used in the task were monomolecular neutral odorants diluted in mineral oil, through which compressed air was bubbled in custom-built olfactometers using mass flow controllers (Aalborg) and solenoid valves (Lee Co). Odor identities were randomized across animals. Odors were obtained at >98% purity from Sigma-Aldrich. Odors and concentrations used were as follows: isoamyl acetate (1:10), pinene (1:5), benzaldehyde (1:10), limonene (1:5), ethyl butyrate (1:20), acetophenone (1:5), octanal (1:5), cis-3-hexen-1-ol (1:10).

Behavior hardware (tone buzzers, port entry beam breaks, odor control valves) was controlled by Bpod microcontroller systems (Sanworks) and a Matlab interface as well as custom Matlab code. Behavioral timestamps and imaging frame acquisition were synchronized using external data acquisition systems (Ni-DAQ). Licks were measured using capacitance sensors (Phidgets) at each port's lick spout.

Behavioral Training

Mice were trained to perform the information seeking task over several weeks of behavioral shaping progressing through the following stages.

1. **Center Training**. Mice received water in the center port (nosepoke) immediately upon triggering it. 1 day, ~15 minutes.

2. **Covered Side Training**. Mice performed alternating blocks of trials where they had to trigger the center port and then enter the left or right port where they received water reward. The incorrect port was covered to prevent entry for the duration of each block of ~50 trials (200uL total reward) each. Trials provided 1-2 drops of 4uL of water with the two sides being equal. The center port odors that would direct mice to each side on later forced trials were presented for the duration of the mouse's triggering of the center port, and the duration of the center poke required to trigger the start of the trial was gradually increased to 200ms. Delivery of water at the side port was delayed up to 1s with gradual increases as mice became proficient at this training stage. The 3s inter-trial interval was also introduced to increase the animal's focus on the task and decrease repetitive behavior. Mice progressed to the next training stage when they achieved >50% complete trial initiations (remaining for the full 200ms of center port odor presentation) and had rapid reaction times (<2s).

3. **Uncovered Side Training**. This stage was not always included but was implemented for animals that had not fully learned the association between center port odors and the required forced side location. Mice completed blocks of ~20-50 trials directed to the left or right side by the odor provided in the center port with the incorrect side uncovered and accessible. They only received water if they chose the correct side port.

4. **Delay Training**. Mice next performed trials with all ports uncovered and pseudorandomly alternated forced left and right trials with odors directing them to the left or the right. All correctly chosen trials were rewarded (1-2 4uL drops) equally on the two sides. The delay between the completion of trial initiation at the center port and water reward delivery following correct choice of the side port was gradually increased by 400-1000ms approximately every 50 trials until the full delay value of 10s was reached. Delays were increased manually while observing mouse behavior to moderate task difficulty and ensure animals maintained adequate motivation. During this stage, a reaction time requirement was instituted such that mice must choose the correct port within a certain amount of time to receive reward. This time was gradually decreased to the final value of 1.2s. Mice moved on to the next training stage when they achieve >70% correct performance at the full delay value.

5. **Introduction of information**. In the next stage, trial rewards became probabilistic and side port odors were introduced. Reward probably was decreased to 50% on all trials. 200ms presentations of the side port odors (A,B,C, and D) occurred 1.2s after the go cue in the correctly chosen side port with the appropriate contingencies, such that odor A was always followed by water reward, odor B was never rewarded, and odors C and D were each followed by water reward on 50% of trials. This allowed mice to begin learning that A and B resolved the trial's outcome and provided information, while C and D did not provide information. 2-3 sessions were performed at 50% reward and then the reward probability was dropped to 25%. Licking responses and the mouse's presence in the reward port were monitored to ensure they were learning the meanings of the side port odors (mice tended to

leave the side port following receipt of odor B). This training stage lasted 5-7 sessions to ensure all mice learned that information was provided.

6. **Choice Training**. Mice were explicitly taught that the Choice center port odor allowed them to receive equal probability water reward at either side port in this stage of training. Task event times, side port odors, and reward probabilities were the same as in the previous stage. Mice performed blocks of 10-50 trials with one side port covered to prevent accessing it. Trials with the appropriate Forced odor (Information or No Information) for the uncovered side were alternated with trials in which the new Choice odor was presented at the center port. Blocks were alternated to ensure mice experienced equal reward rates and total reward amounts following the Choice odor on the left and right side to minimize side bias. This stage lasted for 5-7 sessions, or ~250-300 Choice trials on each side.

7. **Full Task/Preference Measurement**. Following choice training, mice were tested for preference for information on the full version of the task. Both side ports were uncovered, and mice performed all three trial types, Forced Information, Forced No Information, and Choice, pseudorandomly interleaved throughout each session. Trial types were set up pseudorandomly in blocks of 12 to ensure mice performed roughly equal numbers of trials of each type within a session and rewards were assigned in blocks of 8 since trials were rewarded with 25% probability to ensure mice were rewarded frequently enough to maintain their motivation to engage in the task. In initial behavioral tests (n=14), preference was tested for 3 sessions, then the side identities were reversed three times, such that preference was tested for 3 days with information on side 1, 3 days on side 2, 3 days on side 1, and three days on side 2. In mice that were imaged (n=7), preference was tested for 6 days on each side to ensure stabilized learning of the relevant neural representations. On side reversals, the side locations of the information and no information side port odors were switched while center port odors still indicated movement to the same side. This meant that the center port odor that had previously signaled Information on the left side and was followed by odor A or odor B now signaled No Information on the left side and was followed by odor C or odor D. Similarly, mice had to learn that Information had moved from the right to the left side or vice versa.

Water and Delay Titration Experiments

Following side reversals, information remained on the initial side (left or right) and mice were trained for several more sessions (~3) until their preference re-stabilized. Then a stairstep procedure was used to determine the willingness of mice to sacrifice water reward for information. The reward probability on both sides was raised to 50%, but the reward amount on the information side was changed for 3 sessions at a time. Reward values of 4-ul drops were 1 drop, 6 drops, 2 drops, 5 drops, 3 drops, 4 drops, and then these amounts were re-tested in reverse order. Preference was calculated as the mean across the final session in each block at a given reward amount, such that two sessions were used to calculate the preference at each reward amount. A similar procedure, but for single blocks of 6 days at each value, was used for imaged mice.

Information preference was measured across different durations of the delay between side odor presentation and water reward using a similar procedure in which preference was tested for 6 days at each delay value: 1s, 10s, 4s, 10s, 6s. Preference was determined as the mean preference for information in the last two sessions at each delay value.

Stereotactic Surgery

Mice were anesthetized with ketamine (100 mg/kg) and xylazine (10mg/kg) through intraperitoneal injection and received analgesia via buprenorphine SR subcutaneous injection (0.5-1.0mg/kg), had fur shaved from their head, and then were placed in a stereotactic frame. Body temperature was maintained using a heating pad attached to a temperature controller. For lens implantation experiments, a 1-1.5mm round craniotomy centered on the implantation coordinates was made using a dental drill. Dura was removed, and 0.3uL GCaMP6f virus (AAV1.CaMK2a.GCaMP6f.WPRE.bGHpA , Inscopix) was injected into lOFC (ML: 1.0; AP: 2.4; DV: 2.45mm from Bregma) at ~50nL/min using a pulled micropipette. After allowing the virus to diffuse undisturbed for 5min, the needle was removed and a 0.5mm or 1mm diameter and 4.0 mm length microendoscopic GRIN lens with integrated baseplate (Inscopix) was then inserted at a depth just above the injection site and centered over it (ML: 1.0; AP: 2.4; DV: 2.4mm from Bregma). The combined lens and microscope baseplate assembly (Inscopix) was secured to the skull using Metabond (Parkell) and further secured and covered with dental cement (Ortho Jet). Mice recovered for at least 1 week before the beginning of water restriction and experiments.

Histology

Mice were euthanized after anesthesia with ketamine/xylazine by perfusion with 4% paraformaldehyde. Brain tissue was removed for 24hr fixation, and coronal sections (120 um) were cut on a vibratome (Leica). The sections were incubated with far-red neurotrace (640/660, Thermo Fisher Scientific) to label neuronal cell bodies. Images were collected using a Zeiss LSM-710 confocal microscope system. Histology was performed to confirm locations of implanted lenses, as well as expression levels for GCaMP.

Imaging Experiments

OFC activity was recorded during behavioral sessions throughout the course of the information seeking task, including training sessions as described above. Prior to being placed in the behavior chamber, mice had the miniature microscope (Inscopix) attached securely to their skull baseplate. OFC activity was then recorded throughout the behavior session, and the microscope was removed and the baseplate cover reattached following training. Images were recorded with the lowest possible LED illumination to visualize calcium transients with settings largely preserved from session-to-session within each animal. The Z-focus of the microscope was adjusted rarely to maintain the same field of view based on landmarks such as blood vessels.

Quantification and Statistical Analysis

Behavior analysis

Behavior was recorded as nosepokes in the three ports via infrared beam breaks and licks of the reward spout using the Bpod hardware. Behavior was analyzed using custom Matlab scripts. Mice performed approximately 150 trials per session. Initial information preference was determined across up to 300 Choice trials prior to reversing the location of the information from left to right or vice versa. Behavioral preferences across reversals was determined during either each session of the reversal (timecourse plot, Fig.1c) or on the last session tested with information on a given side (overall plot, Fig.S1b). Logistic regression was used to compute the influence of information vs. side bias on preference (Fig.1d). Reaction time was computed as the time elapsed from the go cue until the first entry of the correctly chosen side port.

Behavioral Models

Cognitive decision models, fit to single trial level-choice data, have been used in neuroscience and behavioral economics for decades to provide evidence that a specific type of computation may be underlying choice trends seen in a given dataset[79]. The two main questions we sought to answer with decision model frameworks here were (1) “why do information-preferring mice not always choose information?”, and (2) “can the observed experimental behavioral trends be explained by a model that uses separate but interacting reinforcement learning (RL) machinery for processing traditional water rewards and intrinsic information rewards?”. We hypothesized that there is an interaction between RL-like computations of information value and water value following recent literature[7], and used a list of relevant models of increasing complexity to test this hypothesis. Importantly, the models and experiments in this work do not address “why” mice prefer information (e.g. uncertainty aversion, emotional response to informative cues, etc.), which is an important direction for future dual experimental and theoretical work using this new model organism for information value studies.

In decision model literature, there are a large variety of RL-related models, often becoming quite complex which risks over-parametrization and over-fitting if the number of agents fit and task complexity do not scale with the model complexity[80,81]. Thus, for the purposes and constraints of this study, we chose to use the simplest implementation of RL computations, the Rescorla-Wagner model, following similar recent related work in human models[51]. The models tested are outlined below, and incrementally built up to the full model that contains the minimal necessary components to explain both the water-tuning and delay-tuning experiments (Fig1h,i,FigS2).

First, all models tested transform a decision variable *DV* term, which summarizes the core computation, into a sigmoid-transformed binary choice probability $P(t)_{info}$, here the probability of choosing information at trial *t*:

$$P(t)_{info} = \frac{1}{1 + e^{[-DV(t)]}}$$

Equation 1

The term $P(t)_{info}$ is optimized through a negative loglikelihood minimization:

$$NLL = -\sum_{t=1}^{nTrials} \left\{log\left(P(t)_{info}\right) \cdot C(t)_{info} + log\left(1 - P(t)_{info}\right) \cdot \left(1 - C(t)_{info}\right)\right\}$$

Equation 2

Where is $C(t)_{info}$ = 1 for an info choice and = 0 for a non-info choice. Models were each run with 16 initializations of different starting parameter values, and numerically solved using the *fminsearchbnd* in Matlab using a grid search of parameter windows. The nature of *DV(t)* changes in each model, as summarized below:

RLwater Model:

$$DV(t) = \frac{1}{\sigma} \cdot \{\beta \cdot \Delta Q(t) + \varepsilon\}$$

Equation 3

The time-independent fit parameters are σ, the normalization constant (inverse temperature), β, a relative weight factor for the traditional Rescorla-Wagner RL term of water $\Delta Q(t)$, and ε, a bias term, which allows for a flat baseline tendency towards or away from choosing information that the water term is weighted against. The time-dependent term $\Delta Q(t)$ is defined as the difference of the value functions for the info and no-info sides of the task as usual in binary choice tasks:

$$\Delta Q(t) = Q(t)_{info} - Q(t)_{no-info}$$

Equation 4

Each $Q(t)_i$ for *i=info, no-info* is defined according to the Rescorla-Wagner equation with learning rate α_1 and prediction error $\delta(t)$ when choice *i* is chosen

$$Q(t)_i = Q(t-1)_i + \alpha_1 \cdot \delta(t)_i$$

Equation 5

$$\delta(t)_i = Reward(t-1)_i - Q(t-1)_i$$

Equation 6

When a choice is not chosen, with unchosen choice *j*, then $Q(t)_j = Q(t-1)_j$.

The next model tested contains only information-related terms with a new Rescorla-Wagner term $S(t)$ for information value with information prediction error $\theta(t)$ and learning rate α_2, following analogous logic to recent non-traditional RL equations such as choice prediction errors[82]:

λ-RLinfo Model:

$$DV(t) = \frac{1}{\sigma} \cdot \{S(t) + \varepsilon\}$$

$$S(t) = S(t-1) + \alpha_2 \cdot \theta(t)$$

$$\theta(t) = \left\{\left(\frac{Delay(t-1)}{10\,sec}\right)^{\lambda} - S(t-1)\right\} if\ info\ is\ chosen$$

$$\theta(t) = \{0 - S(t-1)\}\ if\ noinfo\ is\ chosen$$

Equation 6

Thus, information's "value" is $\left(\frac{10\ sec)}{10\ sec}\right)^{\lambda} = 1$ for the maximum delay of 10 seconds, but then is exponentially discounted as delay decreases, following the strong tendency observed in the mice data. When the mice do not choose information, they receive no information value and thus the info-reward is set to zero.

The next model tries to remove the *λ* decay exponent, but combine the *RLwater* term with the *RLinfo* term simplified to have *λ=0,* so that information value = 1 on information choices regardless of delay length.

RLwater + RLinfo Model

$$DV(t) = \frac{1}{\sigma} \cdot \{\beta \cdot \Delta Q(t) + S(t)_{\lambda=0} + \varepsilon\}$$

Equation 7

Then, the full model combining both water value and information value computations is defined below, conceptually motivated by recent work finding both types of value are processed in an interacting way[7].

RLwater + λ-RLinfo Model

$$DV(t) = \frac{1}{\sigma} \cdot \{\beta \cdot \Delta Q(t) + S(t) + \varepsilon\}$$

Equation 8

Last, as a conceptually alternative decision model, we made a weighted win-stay lose-shift (WS-LS) model, that regressed the past trial information and reward conditions to predict the current choice:

WS-LS Model:

$$DV(t) = \frac{1}{\sigma} \cdot \{\beta_1 \cdot InfoRew(t-1) + \beta_2 \cdot NoInfoRew(t-1) + \beta_3 \cdot InfoNoRew(t-1) + \beta_4 \cdot NoInfoNoRew(t-1) + \varepsilon\}$$

Equation 9

Each β_i condition variable is either 0 or 1. For example, *InfoNoRew* = 1 if the past trial was an unrewarded information choice (either forced or choice), but = 0 else.

The total number of mice fit was *N*=10 for all models, but for 4/10 mice only the water tuning data was fit, not the delay tuning data since those mice did not perform delay-tuned task variants. These mice have had their λ discount exponent removed as shown in the Supplementary Table 4 of fit parameters. Then 3/10 mice did not have water tuning data fit. These mice are marked in Supplementary Table 4, but no fit parameters were omitted.

To see how much each RL component improved the full model fit, the average Akaike information criterion (AIC) was calculated according to the standard equation for each model with *k* as the number of free parameters[83]:

$$AIC = 2 \cdot k + 2 \cdot NLL$$

Equation 10

The full ***RLwater + λ-RLinfo*** model is presented in Figure1. The auxiliary models are presented in Figure S2. It is visually clear that only the ***RLwater + λ-RLinfo*** can simultaneously capture the important trends from the data in both water tuning and delay tuning experiments. Future work seeking to do more advanced decision modeling with further corrections to the Rescorla-Wagner formulation should increase the sample size and tune a wider range of experimental conditions, but the current models implemented show that the minimally simple RL formulation that accounts for both water value and information value is able to explain the full range of experimental conditions tested here. As a caveat, we do not claim our fitted parameters are unique or will work for all future studies, a common issue in cognitive decision models because model parameter magnitudes and scaling are not independent and thus degenerate solutions likely exist[79,80]. However, this well-known issue does not undermine the finding that the minimal two-value-type RL model machinery performs well in our context conceptually.

Imaging Data Processing

Seven mice were imaged throughout the course of training and testing in the task. Recordings of GCaMP activity were made using Inscopix miniaturized microscopes (nVista2.0), with uncompressed videos saved using the Inscopix system. Raw video data were spatially downsampled at 4x and saved as Tiff files using the Inscopix Python API. Motion correction was performed within each session using the NormCorre algorithm[84]. However, prior to applying NormCorre, videos were filtered using a difference of Gaussians to extract potential stationary features smaller than cell diameter from the larger background fluctuation. This filter was used to compute a mean template for each video, and the filtered video was motion corrected to the filter template, then calculated shifts were applied to the original unfiltered images. This process was repeated using the filtered template from the central day within preference testing for each mouse, such that the video from each session was motion corrected to a single template for each animal. Cell location footprints and activity were then extracted using the Matlab implementation of CNMF-E[85].

CNMF-E default parameters were used, except that min_corr and min_pnr were adjusted for each animal to maximize cells identified during the initialized step of the algorithm. Cells were filtered automatically using thresholds for temporal sparseness, minimum peak-to-noise, and a constant baseline of activity to remove false positive cell identifications. Then all cells were manually inspected to ensure appropriate neuron-like shape and calcium traces. Using this procedure we were able to register cell spatial footprints across sessions several weeks apart, over the entire months-long course of training, for up to 4 sessions simultaneously using the cell_reg function from the Matlab implementation of CaImAn[57]. We used the calcium signal, C, output by CNMF-E that is smoothed with a kernel based on the timecourse of GCaMP6 signal decay as the calcium activity of each cell across each session. Because CNMF-E subtracts the background signal from each frame individually, the output activity C may be considered a scaled version of the change in fluorescence over background ($\Delta F/F$) at each frame. Moreover, given this background subtraction, this signal is already normalized. We therefore analyzed and report simple “calcium activity.” However, a standard normalization procedure computing a z-score for each cell’s activity across each session was tested and did not affect the results.

Activity in each session was aligned to behavioral events using DAQ timestamps, and we analyzed neural activity across all available mice considering cells across animals as a single pool for single-cell analyses. Our CEBRA population analyses comparing model fits across animals support this approach. The main dataset consisted of cells registered across 4 sessions surrounding a side reversal such that there were 2 sessions with information on each side (left and right), with 56-216 cells per animal across 7 animals. Datasets for delay comparisons and learning also consisted of 4 sessions per animal across 6 and 7 animals, respectively, with similar numbers of cells per animal.

Analysis Window: pre- and post-event activity

We determined a 1-second pre- and post-event activity window for neural analysis both by observing the raw odor responses in our task and because, as determined by tests with a photoionization detector, the odor stimulus did not reach the animal in the nosepoke port until 0.2s after odor valve opening, which was the event detected by the imaging hardware. Thus, we have used 1.2-0.2s before events, and 0.2-1.2s after events, as the pre- and post-event time periods for our analysis.

Heatmaps

Calcium activity population heatmap plots display the mean activity for each cell across trials in each condition, with the cell’s mean activity in that condition across the pre-event 1s period subtracted. Difference heatmaps show differences in mean activity between conditions without pre-event mean subtraction. Plots within panels are all sorted by the indicated activity.

Cells with significant conditional activity differences (differential activity)

Cells with significantly different activity between two conditions were determined using a bootstrapping procedure. For each cell, we shuffled trials randomly between the two conditions 1000 times. On each shuffle, we calculated the difference between the cell’s mean activity in each of the two conditions, both at each frame in the analysis window and the mean across the pre- and post-event periods. Cells that had significantly different activity between the two conditions were then defined as cells that had an observed absolute value of the difference in mean activity between the conditions in the post-event period that was greater than 95% of the shuffled data and an absolute difference greater than a threshold of 0.1.

Cells with significant conditional responses to task events (responding cells)

Cells were identified as responding to an event in a task condition using a rank sum test to determine if their activity in the post-event period in that condition was different from activity in the pre-event period in that condition, which we considered the baseline activity for that event. To accommodate baseline activity that was already elevated due to earlier trial events given the close proximity in time of some stimuli in the task, we determined the maximum activity for each cell on each trial in the pre-event period and fit an exponential decay function based on GCaMP6f fluorescence (λ=4) to determine its predicted activity in the post-stimulus period. If the absolute value of difference between the mean of that predicted activity in the post-event period and the mean of the observed activity in the post-event period on that trial was greater than 0.2, the GCaMP fluorescence-predicted activity in the post-event period was used as the baseline. If not, the pre-event activity served as the baseline. Cells with significant responses were then identified as those in that condition with mean activity in the post-event period significantly greater than baseline using a rank-sum test and with the value of that difference greater than 0.1.

Mean activity

Mean activity was calculated across either all cells or cells within a particular sub-population as indicated. Each cell's mean activity on the indicated trial type was calculated, and then all cell's mean activities were averaged.

Mean absolute difference in activity

For the indicated conditions, each cell's mean activity in each condition was calculated as above. The absolute value of the difference in these condition means was computed for each cell, and then values were averaged across the population. We also computed a cross-validated absolute mean differential activity, in which we took the mean of each cell's activity separately for half the trials within each condition, computed the difference between the two conditions for these two trial group means separately, and then averaged these differences between the two groups, using the sign from that cell's difference in the other group. This eliminated the positive bias from the absolute average activity difference but did not change our results. We therefore show the more straightforward mean absolute difference in activity.

Decoding

Linear classifiers (support vector machines) were used to decode trial types. Classifiers (Matlab function fitclinear, default support vector machine) were trained on 80% of data and then tested on 20% of held out data. Trials were balanced between relevant types. Classification was performed on 200ms bins of neural data, with calcium activity averaged for each cell on each trial within each bin.

PCA of information-no information axis

To compute an axis along the information-no information dimension in population activity space, we calculated the difference between the mean activity on information forced and the mean on no information forced trials for each cell following the center port odor presentation (0.2-1.2s after odor valve opening). We then computed the principal components of that population activity (Matlab function svd). The first component captured >80% of the variance in this activity, and we therefore projected the population activity in each trial, and the mean within each condition (information and no information) onto that component.

CEBRA Modeling

We used Consistent EmBeddings of high-dimensional Recordings using Auxiliary variables (CEBRA) to better understand how a complex task such as ours is represented at the neural level. The complexity of our task, together with the possibility that the OFC, as a frontal region, contains neural representations that are more compressed and processed compared to other regions, suggested that neural task representations are more likely to have arisen from the non-linear combination of neural signals[35,38]. A nonlinear latent variable model such as CEBRA, unlike principal components analysis (PCA), for example, reduces the chance that an analysis is fitting task-irrelevant noise or missing important nonlinear signals that are crucial to accurately capture how a task is encoded in the circuit.

More specifically, we used CEBRA to obtain low-dimensional task embeddings by jointly fitting neural data and task-relevant behavioral variables in a time-dependent manner. We generated time-stamped labels expressing task structure and mouse behavior, as further specified below, and fitted these together with the neural data by minimizing an InfoNCE contrastive loss objective[58]. We fitted the model to the entire cohort of animals at once, such that contrastive samples were chosen from across all animals. We thus obtained latent embeddings that were informative and consistent across animals. We examined the geometry of the information and water value representations in the embeddings and measured the differences between these two representations over the course of each trial. We

confirmed that the embeddings were informative by decoding task-relevant variables and comparing results to shuffled versions which scrambled the association of particular labels with the recordings. We showed that they were consistent by examining differences between information and water value representations in the embedding space across recording sessions and animals.

We thus derived joint low-dimensional embeddings of high-dimensional neural and behavioral data while also capturing time-dependent information using CEBRA. We fit CEBRA models with the following specifications: model architecture = “offset1-model”, batch size = 1024, distance = ‘cosine’, conditional = ‘time_delta’, temperature = ‘auto’, learning rate = 0.001, max iterations = 5000, output dimension = 3, number hidden units = 100. We arrived at these settings by fitting CEBRA models on the data from 3 mice and picking the settings that achieved the lowest overall InfoNCE[86] loss value after 10000 iterations; during this process we observed that only half the iterations were necessary for the loss to converge so we set max iterations = 5000 to minimize compute.

We fit two types of models which we will refer to as full and delay henceforth. The full model is fit on data with a long delay (t_d_l = 10 seconds) between odor 2 and reward delivery, while the delay model is fit on data with both short (t_d_s = 1 second) and long delay (t_d_l = 10 seconds) and is aimed at embedding the two delay types in a joint space for comparison.

The full model was fit on 6 mice (JB432, JB426, JB434, JB413, JB424, JB425) with cell and trial counts as specified below (Table 1). Importantly, we employed a multi-session setup in which each session with its unique dimensionality was added into the fitting process accompanied by shared context labels (Table 3) which allowed CEBRA to learn a shared embedding space for signals from all sessions (across all mice). We used sessions from our main data set balanced before and after information side reversal, such that each mouse was fit with two sessions with information on each side (left/right). In order to achieve an unbiased representation across all conditions we equilibrated the number of trials across conditions such that there were equal numbers of trials across the four conditions resulting from the interaction of “reward vs. no reward” and “information vs. no information”. For the full model, the number of time points per trial is fixed at n_time = 320 time points per trial throughout, ranging from the initiation of the last (if multiple) presentation of the center port odor all the way until after reward delivery.

Mouse	Session	Number of cells	Number of trials (per session)
JB432	20220526	442	56
JB432	20220527	522	44
JB432	20220613	505	68
JB432	20220614	426	54
JB426	20220302	314	48
JB426	20220303	599	52
JB426	20220316	381	32
JB426	20220317	312	56
JB434	20220526	417	24
JB434	20220527	323	60
JB434	20220606	547	16
JB434	20220607	516	64
JB413	20211123	331	48
JB413	20211124	475	24
JB413	20211220	397	44
JB413	20211223	334	68
JB424	20220210	414	48
JB424	20220211	337	32
JB424	20220223	416	24
JB424	20220224	358	32
JB425	20220203	292	36
JB425	20220207	341	36

JB425	20220217	177	48
JB425	20220218	179	56

Table 1: Cells per session for *full* CEBRA model

The *delay* model was fit on 6 mice (JB432, JB433, JB434, JB426, JB424, JB425) with cell and trial counts (Table 2) and label types (Table 3) with n_time=113, as specified below. In order to make the two delay types (*short* and *long* delay) comparable, with identical context labels across sessions allowing sessions with variable cell counts to be fitted together in one model, we cut out the outcome period in the *long* delay trials and concatenated it after the first ~1s of the 10s delay period (omitting most of the longer delay) thus making the overall duration of short and long delay trials the same. These sessions do not include a reversal of the information port location. As before, in order to achieve an unbiased representation across all conditions we equilibrated the number of trials across conditions such that there were equal numbers of trials across the four conditions resulting from the interaction of "reward vs. no reward" and "information vs. no information"; in case that resulted in less than 1 trial per condition, we equilibrated instead by "information vs. no information" only (session marked with * in Table 2).

Unlike the *full* model, we matched cells across all recording sessions for a particular mouse before fitting the *delay* model (hence the matching cell counts in Table 2). We then combined all of a mouse's recording sessions into a single array and fitted the recordings and labels (Table 3) with CEBRA in a *single*-session setup. This allowed us to derive a common embedding space across both delay types.

Mouse	Session	Number of cells	Number of trials
JB432	20220623	258	44
JB432	20220627	258	40
JB432	20220706	258	92
JB432	20220707	258	56
JB433	20220629	136	64
JB433	20220630	136	72
JB433	20220708	136	64
JB433	20220711	136	64
JB426	20220616	72	56
JB426	20220617	72	44
JB426	20220711	72	92
JB426	20220713	72	88
JB434	20220812	223	36
JB434	20220817	223	24
JB434	20220824	223	203*
JB434	20220825	223	84
JB424	20220527	169	48
JB424	20220531	169	80
JB424	20220606	169	120
JB424	20220607	169	80
JB425	20220518	135	84
JB425	20220519	135	68
JB425	20220607	135	80
JB425	20220609	135	88

Table 2: Cells per session for *delay* CEBRA model

We fit all CEBRA models on the neural data (as defined above) together with the following label types (Table 3). Labels are zero everywhere except between the start and end frame where they are set to nonzero values depending on label type.

Label Type	Range	Type	Start Frame	End Frame
trial time index	[1, 2,3,…,n_time]	discrete	1	n_time
rewarded vs unrewarded	[-1, 0, 1]	discrete	274	n_time
info vs no info	[-1, 0, 1]	discrete	40	n_time
info rewarded vs unrewarded	[0, 1, 2]	discrete	70	n_time
no info rewarded vs unrewarded	[-2, -1, 0]	discrete	274	n_time
trial epoch center port odor	[0, 1]	discrete	40	44
trial epoch side port odor	[0, 2]	discrete	70	74
trial epoch outcome	[0, 3]	discrete	74	n_time

Table 3: CEBRA labels

We also obtained the decoding performance for the CEBRA embeddings. We divided the data from each animal and session separately into a train and a test set such that data from each animal and session was present in both sets, with 40% of the data being assigned to the test set. We obtained a CEBRA fit on the train set with the parameters described above (Table 1-3) and then projected the train set recordings into the embedding space. We thus obtained 3-dimensional embeddings for every trial in the train set and used these trajectories to train a k-Nearest Neighbors classifier with a cosine-distance metric and k=20. We fitted separate classifiers for each label type to be decoded. We then used these classifiers to predict a given label type at each point in a trial for all the recordings in the test set. To quantify performance, we calculated the balanced accuracy score across a bin of 5 samples with a step size of 1, and averaged results across sessions, animals and runs. The results we report are averaged across 5 runs, meaning 5 separate CEBRA fits with different random seeds.

We also fitted label-level controls. There were 4 different label-level control types: “shuffled trials all labels”, “shuffled max all labels”, “shuffled info label” and “shuffled reward label”. For the “shuffled trials all labels” condition, we randomly scrambled the assignment of trials to their original time-varying labels, for all the label types, but left the relation between labels and the time they were assigned in a trial (Table 3) intact. For the “shuffled max all labels” condition, we randomly scrambled both the trial- and the time-affiliation of a given label, for all label types, thus resulting in the maximum amount of signal scrambling. For the “shuffled info label” condition, we only scrambled the trial-label assignment for the “info vs no info” label type (Table 3), leaving all others intact. For the “shuffled reward label” condition, we only scrambled the trial-label assignment for the “rewarded vs unrewarded” label (Table 3), leaving all others intact. Otherwise, controls were fitted identically to the full model. Results are again reported as an average across 5 runs, using different random seeds to achieve the train/test allocation (Fig. S9).

In addition, we fit cell-level controls. There were two different cell-level control types, 60% and 20%, with the percentage number indicating how much of the original population of recorded cells was used (Table 1 or 2). So, for instance, in the 60% condition, only 60% of the cells in a given session were randomly picked to be included in the fit. This was also repeated across 5 runs, using 5 different random seeds to subsample the cells. Results are reported averaged across sessions, animals and runs.

We also obtained the distance between individual trajectories in the embedding space. For that we projected all the recordings into the embedding space obtained by the *full* CEBRA model fit. We then calculated the pairwise Euclidean distance between the trial-averaged trajectories of different conditions of interest. In particular, we considered trajectories based on “rewarded vs unrewarded”, “info vs no info”, as well as the interaction of the two label types. We averaged results across sessions, animals, and runs.

Table 4

Mouse	σ	β	α1	α2	InfoVal	λ	ε
1	0.244	0.992	0.302	0.200	1	N/A	-0.258
2	0.335	0.634	0.300	0.208	1	N/A	-0.236
3	0.235	0.600	0.312	0.250	1	N/A	-0.261
4	0.370	0.865	0.302	0.237	1	N/A	-0.267
5	0.303	0.600	0.733	0.433	1	2.000	-0.232
6	0.350	0.600	0.960	0.459	1	1.875	-0.291
7	0.235	0.934	0.303	0.463	1	0.409	-0.252
8	0.227	0.611	0.461	0.106	1	0.300	-0.271
9	0.269	0.824	0.484	0.153	1	0.653	-0.225
10	0.399	0.619	0.829	0.061	1	0.323	-0.232
Mean	0.297	0.728	0.499	0.257	1	0.927	-0.253
SEM	0.020	0.050	0.251	0.146	N/A	0.794	0.021

Table 4: Full behavioral model parameters for each animal cross-animal parameter mean and SEM. See Methods for behavior model details.

Table 5

Model	Mean Accuracy	SEM	AIC/N	SEM
RLwater	0.481	0.054	1.388	0.041
λ-RLinfo	0.720	0.028	1.064	0.058
RLwater + RLinfo	0.713	0.028	1.071	0.043
RLwater + λ-RLinfo	0.734	0.011	1.078	0.026
WS-LS	0.643	0.013	1.339	0.049

Table 5: Behavioral model comparison. Accuracy (SEM) and Aikake Information Criterion normalized by sample size, N, (SEM) reported across animals for each tested behavioral model. See Methods for behavioral model details.

SUPPLEMENTAL INFORMATION

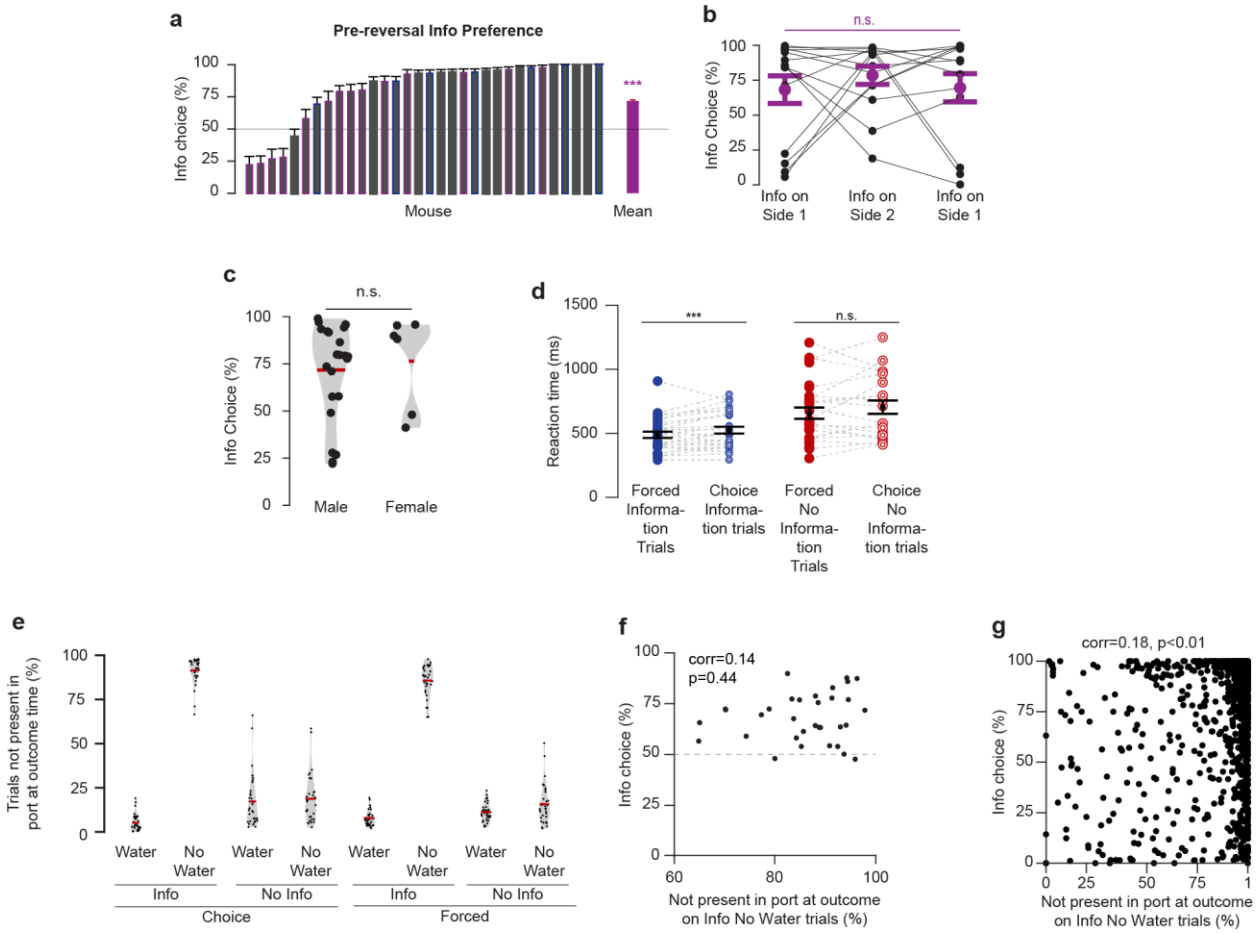

Figure S1: Behavioral task performance

(a) Pre-reversal information preference for all animals. Blue border=imaged animals, purple border=animals in original cohort shown in b. Bars show mean preference for information port on choice trials across last three testing sessions prior to side reversal, +/- 95% confidence interval, p<0.01 sign-rank test. (b) Preference for the current information side on the last session prior to reversal. Points in purple show mean +/- 95% confidence interval across animals, black points and lines show individual animals. Sign-rank test for significance. (c) Overall choice of information as in Fig.1b for male and female mice. Mean information preference for each animal on choice trials in preference testing sessions prior to side reversal (black dots). Red line indicates mean. Difference between males and females n.s. in rank-sum test. (d) Reaction times on forced versus choice trials for all animals. Mean reaction time on trials prior to reversal for the trial types indicated. Single points indicate per-animal mean, black indicates population mean +/-s.e.m. (e) Leaving the side reward port. Plot shows fraction of trials of indicated types in which the mouse was not present in the chosen reward port at the time of the reward outcome (black dots). Red line indicates population mean. (f,g) Low correlation between leaving the port on unrewarded trials and information preference. (f) Leaving on information no water trials vs. information preference. X-axis shows the fraction of trials not present at outcome on information no water trials. y-axis shows information preference. Trials include all preference trials across side reversals. (g) Fraction of trials not present at outcome on information no water trials vs. information preference. Each point represents a single testing session, sessions are shown for all mice.

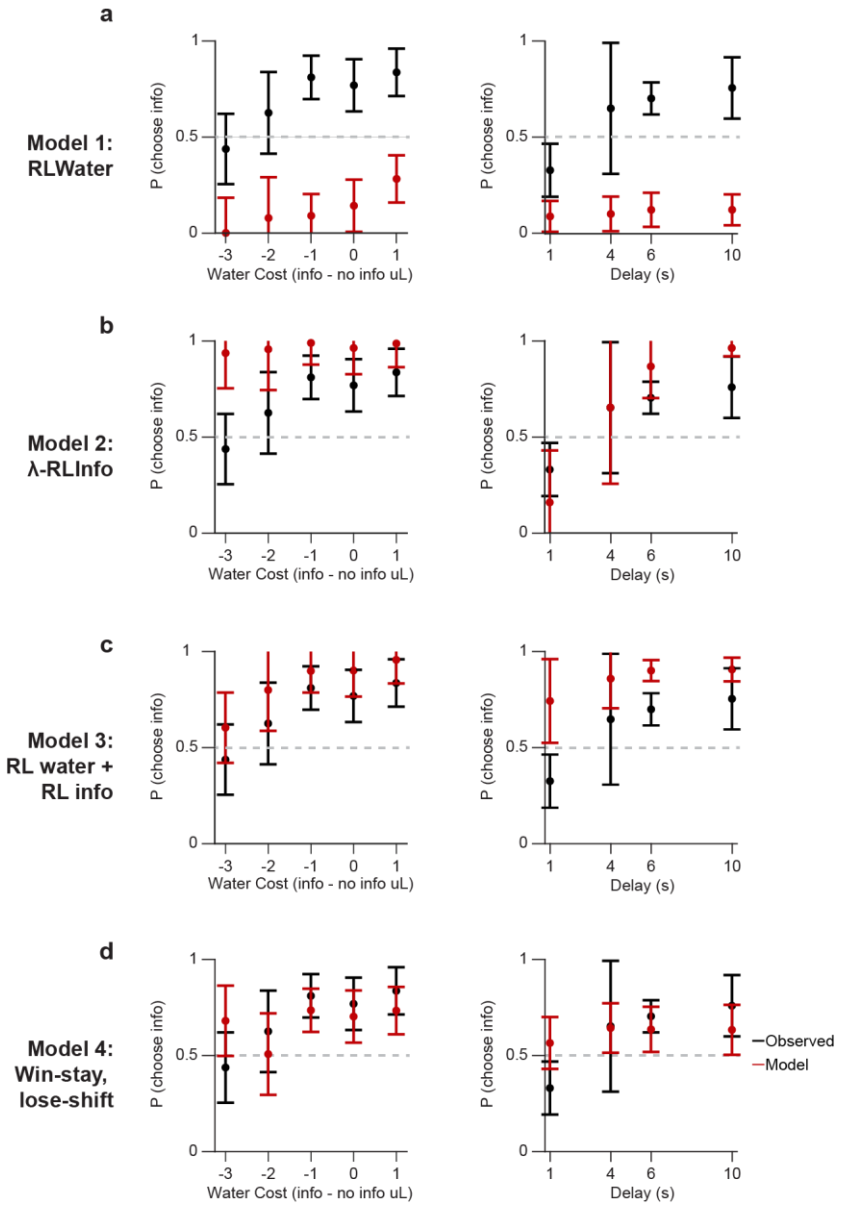

Figure S2: Behavior model details

An ablation study of our decision model in Fig1 and Methods was performed to show the behavioral effects of dropping each of the key model terms and variables. (a) When the information value Rescorla-Wagner term was dropped (RLwater), the ability to predict information preference completely collapsed. The predicted preference dropped below 50% because the model was fit to actual mouse choices and trial variables, and the mice on average preferred information. (b) When the Rescorla-Wagner term of water value was dropped but the delay-discounted Information term was kept (λ-RLinfo), the model overestimated the information preference at the highest delay value and underestimated the Information preference at the smallest delay value. It also completely failed to predict the behavior in the water value-tuning experiments. (c) When the discount exponent was dropped on the information value term, so that the information value was unchanged at all delay durations (RLwater + RLinfo), the model failed to predict the results of the delay tuning experiment. (d) Last, we tested a win-stay, lose-shift model where mice have sigmoid transformed regression coefficients for each possible previous trial type and outcome (info rewarded, no-info rewarded, info unrewarded, no-info unrewarded). For example, mice may want to switch more often after unrewarded trials. This model predicted overall average behavior in a coarse-grained way, but failed to predict behavioral changes as relative water value and delay duration were tuned. Error bars are 95% CI.

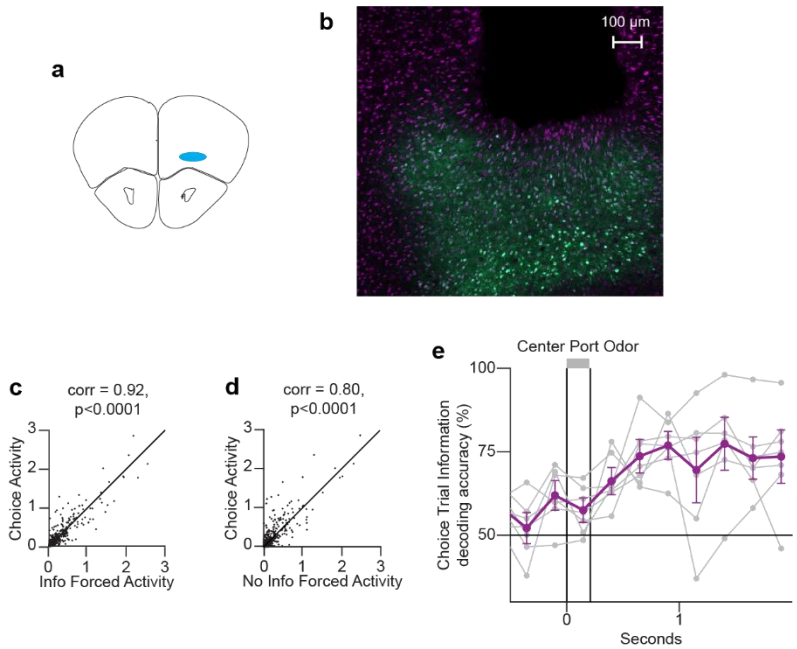

Figure S3: Imaging locations and additional plots related to Figure 2

(a) Schematic of GRIN lens placement in lOFC for imaging. GRIN lenses were implanted at coordinates ML: 1.0; AP: 2.4; DV: 2.4mm from Bregma. (b) GCaMP6F expression below lens placement in OFC at the conclusion of imaging. (c) Correlation between activity in response to center port information forced trial odor and choice trial odor on all choice trials. Each point indicates the mean-subtracted activity in response to odor on the indicated trial types for a single cell. (d) Correlation between activity in response to center port no information forced trial odor and choice trial odor on all choice trials. Each point indicates the mean-subtracted activity in response to odor on the indicated trial types for a single cell. (e) Decoding choice of information vs. no information following center port odor presentation on choice trials. Purple, mean +/-95% CI across mice, grey=individual animals. Trial counts balanced between information and no information.

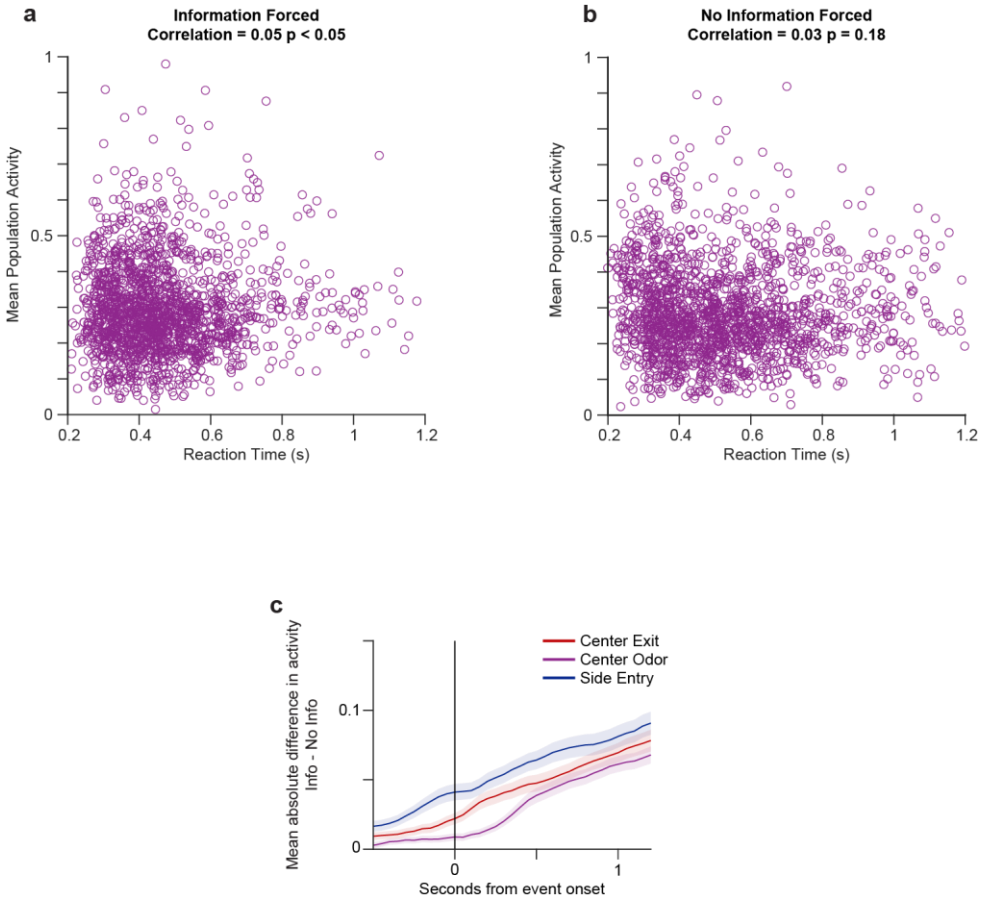

Figure S4: Information representation is not based on movement

(a,b) Low of correlation between reaction time and center port odor neural response. (a) Reaction time on each information forced trial plotted against mean population response to center port odor. (b) Reaction time on each no information forced trial plotted against mean population response to odor. (c) Information CS+ response aligned to odor onset and port entry and exit. Mean absolute difference in information forced – no information forced activity is plotted aligned to the onset of the indicated event.

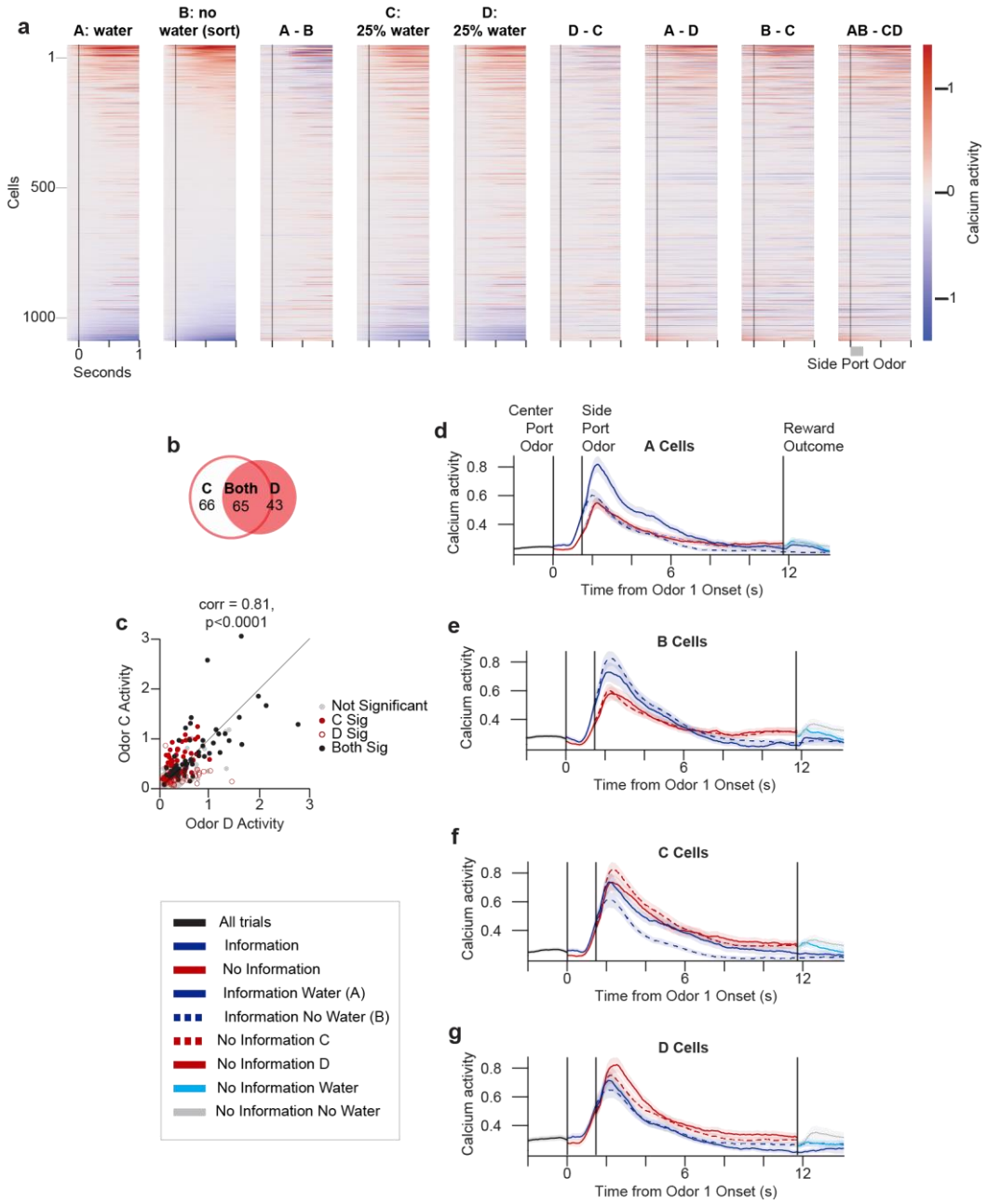

Figure S5: OFC side port odor responses

(a) Population activity of responses to the side port odors and their differences. Each row shows the mean-subtracted activity for each cell across all panels, differences show difference in mean activity. All plots are sorted by the response to odor B. (b) Number of cells responding to side port odors C and D and their overlap. N=1138 total recorded cells. (c) Correlation between activity in response to no information side port odor C and D. Each point indicates the mean-subtracted activity in response to each odor for a single cell. (d-g) Plots show mean activity across the entire trial for each indicated subpopulation of cells. Pre-trial mean activity (1s) is subtracted. Cells were selected based on responding to the indicated odor(s). Plots show the mean activity on trials of each type within that subpopulation.

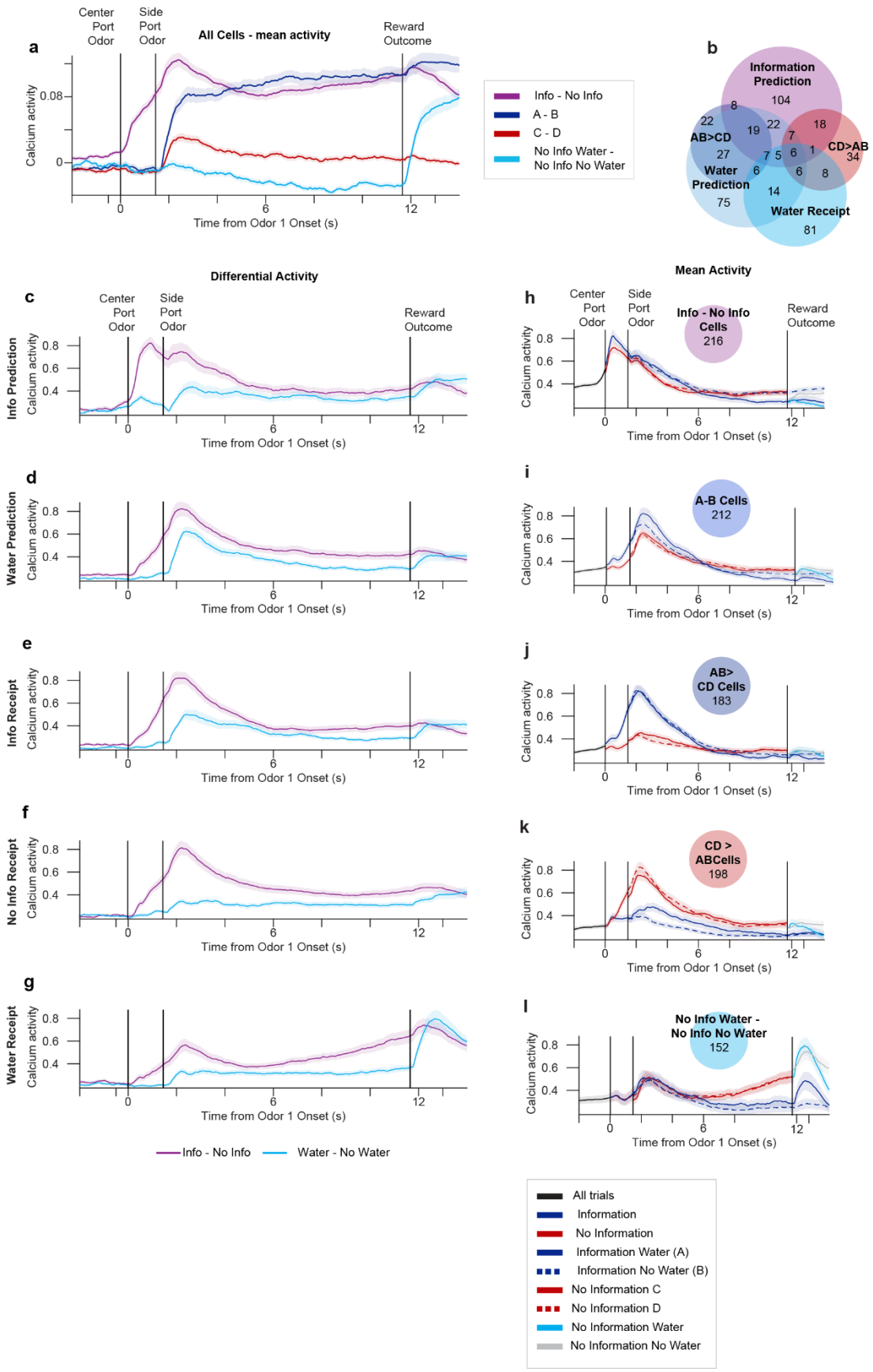
a
All Cells - mean activity
Center Port Odor
Side Port Odor
Reward Outcome
Calcium activity
Time from Odor 1 Onset (s)
Info - No Info
A - B
C - D
No Info Water - No Info No Water
b
Information Prediction
104
AB>CD
27
CD>AB
34
Water Prediction
75
Water Receipt
81
Differential Activity
c
Info Prediction
d
Water Prediction
e
Info Receipt
f
No Info Receipt
g
Water Receipt
Info - No Info
Water - No Water
Mean Activity
h
Info - No Info Cells
216
i
A-B Cells
212
j
AB> CD Cells
183
k
CD > ABCells
198
l
No Info Water - No Info No Water
152
All trials
Information
No Information
Information Water (A)
Information No Water (B)
No Information C
No Information D
No Information Water
No Information No Water

Figure S6: OFC subpopulation representations of information and water

(a) Mean absolute difference in calcium activity between the indicated conditions (mean activity for 1-s pre-trial is subtracted). (b) Number of cells in each of the indicated subpopulations, defined by differential activity between the indicated conditions. N=1138 cells. Information prediction cells show differential responses to the information and no information forced trial center port odors. AB>CD have a positive differential response to side port odors A and B vs C and D. CD>AB have a negative differential response to CD vs AB. Water prediction cells have a differential response to odor A vs odor B. Water receipt cells have a differential response to no information water outcomes vs no information no water outcomes. (c-g) Mean absolute value of the difference in activity between either information and no information (purple) or rewarded and unrewarded (blue) trials of the indicated subpopulation of cells (mean activity 1s pre-trial subtracted). (h-l) Mean activity on trials of each type for the indicated subpopulations (of indicated cell counts, N=1138) (mean activity 1s pre-trial subtracted).

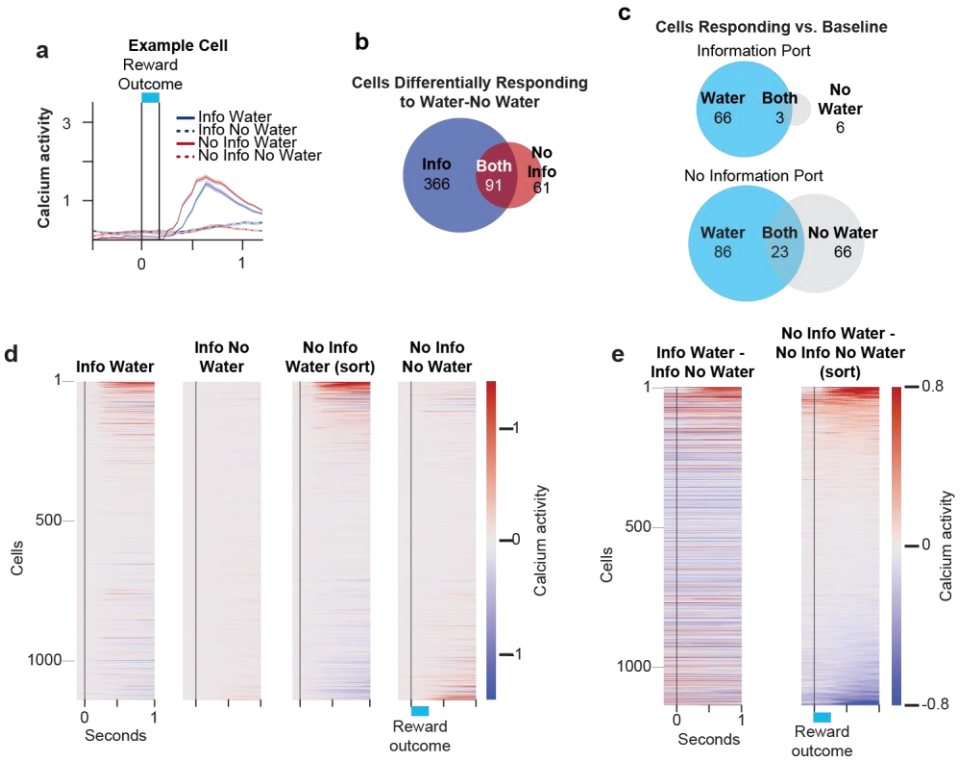

Figure S7: OFC Water US Representation

(a) Responses of a single example cell to water reward outcome. Mean activity in response to water reward or omission at the outcome time at the end of the trial. Mean activity 1-s prior to outcome subtracted for plotting. (b) Cells with differential activity to water reward. Number of cells with differential activity between information rewarded (water) and unrewarded (no water) trials at the outcome period and between no information rewarded and unrewarded trials at the outcome period, and their overlap. N=1138 total recorded cells. (c) Number of cells responding with increased activity at the outcome time to rewarded and unrewarded information trials (top) and rewarded and unrewarded no information trials (bottom) and their overlap. (d) Population activity of responses to water reward or its omission on information (left) and no information (right) trials. Each row shows the mean-subtracted activity for each cell across all the panels. All plots are sorted by the response to water reward on no information trials. (e) Population activity of the difference in responses to water reward or its omission on information (left) and no information (right) trials. Each row shows the activity for each cell across all the panels. Both plots are sorted by the difference in response on no information trials.

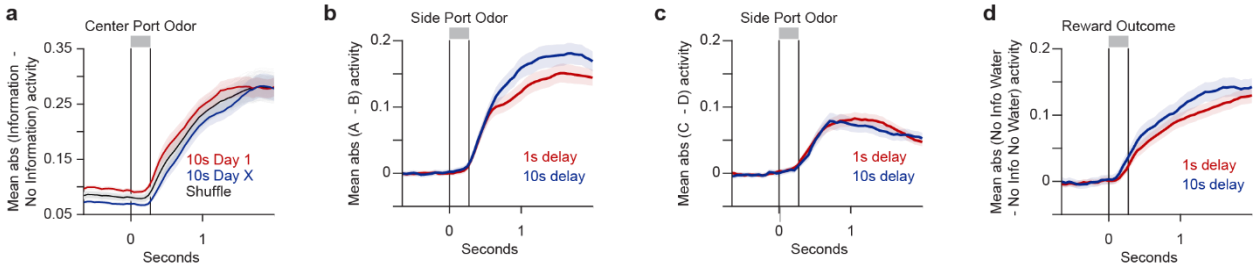

Figure S8: Delay-based information value additional analyses

(a) Comparison of sessions >1 week apart both with 10s delay between side port odor and reward outcome. Mean across the population of the absolute value of the difference between mean activity on information forced trials and mean activity on no information forced trials in each cell. Means calculated separately for sessions on each day and on shuffled activity between the two days. Delay sessions show s.e.m., shuffled data show 1000 different shuffled conditions. (b) Mean water CS+ representation at 10s and 1 s delay. Mean across the population of the absolute value of the difference between mean activity in response to odor A and odor B in each cell. Means +/- s.e.m. calculated separately for sessions with 10s and 1s delay. (c) As in (b) but for no information side port odors C and D. (d) As in (b) but for no information reward vs no reward outcomes.

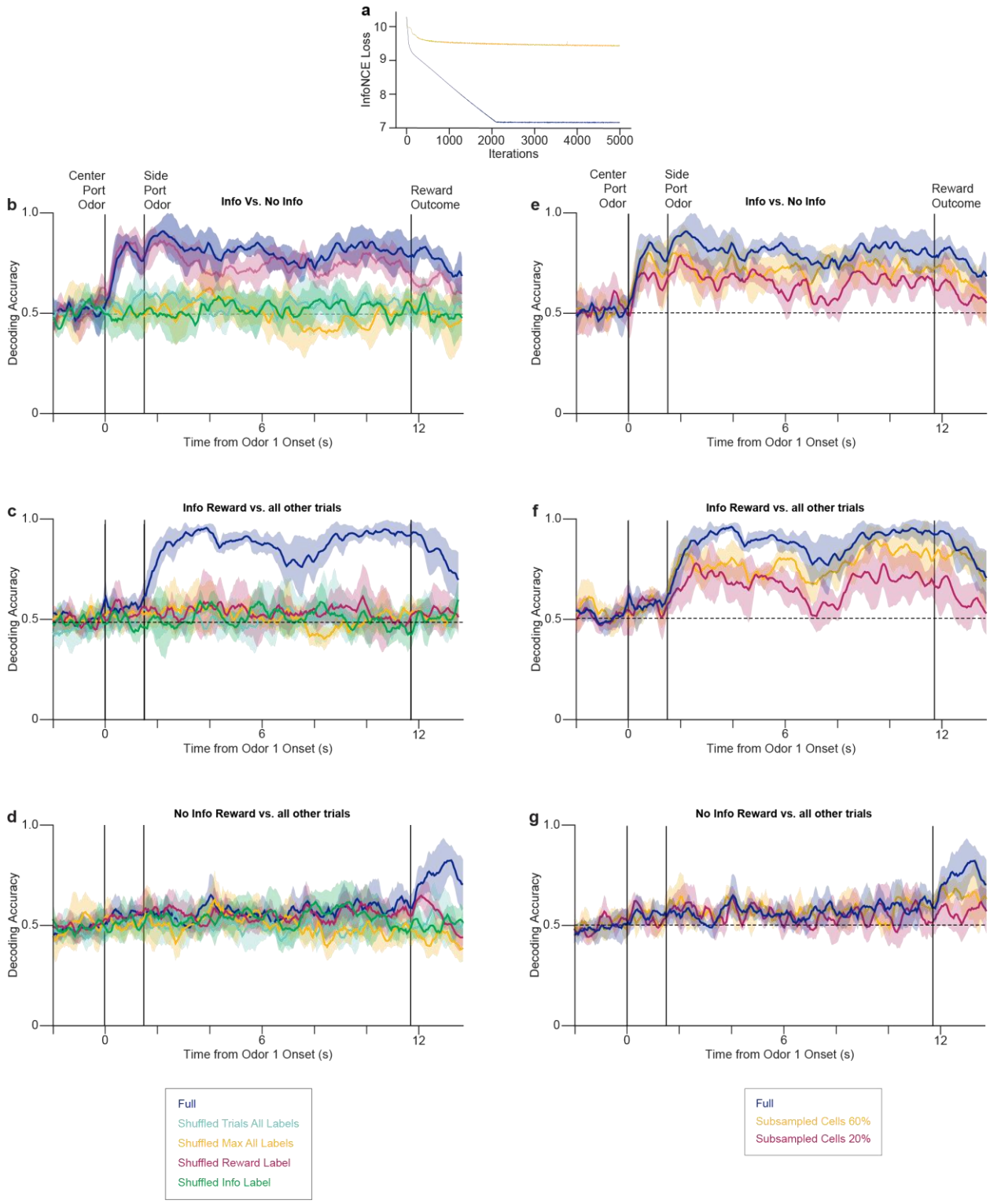

Figure S9: CEBRA supplement

(a) Comparison of the model's contrastive loss as measured by InfoNCE (Info Noise-Contrastive Estimation)[86] with intact versus shuffled labels. Lower loss suggests a better fit model. Decoding accuracy of (b) information trial type versus no information trial type for all trials, (c) presence of reward on information trials and (d) presence of reward on no information trials. In (b-d), the legends are defined as follows. In the "shuffled trials all labels" condition, the assignment of trials was randomized relative to their original time-varying labels, for all the label types, but we left the

matching between labels and the time they were assigned in a trial (Table 3) intact. In the “shuffled max all labels” condition, we randomly scrambled both the trial- and the time-affiliation of a given label, for all label types, thus resulting in the maximum amount of signal scrambling. For the “shuffled info label” condition, we only scrambled the trial-label assignment for the “info vs no info” label type (Table 3), leaving all others intact. For the “shuffled reward label” condition, we only scrambled the trial-label assignment for the “rewarded vs unrewarded” label (Table 3), leaving all others intact. "Full" is the finalized intact labels for the full model. Results are reported as an average across 5 runs, using different random seeds to achieve the train/test allocation. (e-g) Models fit to only random sub-populations of cells were also used as controls to show that embedding quality deteriorated with reduced sample size. 20% (magenta) and 60% (orange) of OFC neurons recorded were sub-sampled and compared to the full sample (blue). Each sampling was the average of five runs, using five different random seeds to subsample the cells. Results are reported averaged across sessions, animals and runs with standard deviation error bars.