Rapid Learning of Odor–Value Association in the Olfactory Striatum

Daniel J. Millman^{1,2} and Venkatesh N. Murthy^{2,3}

¹Graduate Program in Neuroscience, Harvard University, Cambridge, Massachusetts 02138, ²Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, and ³Center for Brain Science, Harvard University, Cambridge, Massachusetts 02138

Rodents can successfully learn multiple novel stimulus–response associations after only a few repetitions when the contingencies predict reward. The circuits modified during such reinforcement learning to support decision-making are not known, but the olfactory tubercle (OT) and posterior piriform cortex (pPC) are candidates for decoding reward category from olfactory sensory input and relaying this information to cognitive and motor areas. Through single-cell recordings in behaving male and female C57BL/6 mice, we show here that an explicit representation for reward category emerges in the OT within minutes of learning a novel odor–reward association, whereas the pPC lacks an explicit representation even after weeks of overtraining. The explicit reward category representation in OT is visible in the first sniff (50–100 ms) of an odor on each trial, and precedes the motor action. Together, these results suggest that the coding of stimulus information required for reward prediction does not occur within olfactory cortex, but rather in circuits involving the olfactory striatum.

Key words: decoding; olfactory tubercle; piriform cortex; rapid learning; reinforcement learning; reward categorization

Significance Statement

Rodents are olfactory specialists and can use odors to learn contingencies quickly and well. We have found that mice can readily learn to place multiple odors into rewarded and unrewarded categories. Once they have learned the rule, they can do such categorization in a matter of minutes (<10 trials). We found that neural activity in olfactory cortex largely reflects sensory coding, with very little explicit information about categories. By contrast, neural activity in a brain region in the ventral striatum is rapidly modified in a matter of minutes to reflect reward category. Our experiments set up a paradigm for studying rapid sensorimotor reinforcement in a circuit that is right at the interface of sensory input and reward areas.

Introduction

Reinforcement learning can quickly improve decision-making by discovering and emphasizing features of the environment (e.g., sensory inputs) and actions of the animal (e.g., motor outputs) that correlate with positive or negative outcomes (Sutton and Barto, 1998; Dayan and Niv, 2008). Rodents can successfully learn to associate odor cues with motor responses following only a few repetitions of trial-and-error per odor when the correct response is reinforced by reward (Slotnick et al., 2000; Slotnick,

2001). The resulting decision process can be highly odor specific: mice can accurately determine the presence or absence of a single rewarded odor hidden within previously unexperienced mixtures of more than a dozen distractors with similar molecular features as the target-rewarded odor (Rokni et al., 2014).

Odor information is first collated in the olfactory bulb (OB), and converges on multiple higher-order brain regions, including the posterior piriform cortex (pPC) and the olfactory tubercle (OT; Ikemoto, 2007; Wesson and Wilson, 2011; Giessel and Datta, 2014; Yamaguchi, 2017; Zhang et al., 2017b). pPC is an association cortex, with projections to multiple cognitive brain regions, including orbitofrontal cortex, amygdala, medial temporal lobe (Johnson et al., 2000; Diodato et al., 2016). OT is located in the ventral striatum (Heimer et al., 1982; Ikemoto, 2007) and is specialized for olfaction, with a well defined cell body layer that receives input directly from OB and olfactory cortical areas (Wesson and Wilson, 2011; Zhang et al., 2017b). The OT is a heterogeneous structure, with different types of neurons, pallidal and striatal subregions, and mesoscopic structures such the islands of Calleja (Heimer et al., 1982; Wesson and Wilson, 2011; Giessel and Datta, 2014). In turn, OT is by far the largest source of olfactory input to the ventral tegmental area (VTA; Watabe-

Received Nov. 2, 2019; revised Mar. 30, 2020; accepted Apr. 6, 2020.

Author contributions: D.J.M. and V.N.M. designed research; D.J.M. performed research; D.J.M. and V.N.M. analyzed data; D.J.M. and V.N.M. wrote the paper.

This work was supported, in part, by a grant from the National Institutes of Health (R01-DC-017311) to V.N.M. D.J.M. was supported by a National Science Foundation Graduate Research Fellowship under Grant DGE1144152 and National Research Service Award F31-DC-014602 from the National Institute on Deafness and Other Communication Disorders. Microscopy on fixed tissue samples was performed at the Harvard Center for Biological Imaging. We thank Nuné Martiros, Nao Uchida, and Kenneth Blum for critical feedback that significantly improved the manuscript.

The authors declare no competing financial interests.

D.J. Millman's present address: Allen Institute for Brain Science, Seattle, WA 98109.

Correspondence should be addressed to Venkatesh N. Murthy at vmurthy@fas.harvard.edu.

<https://doi.org/10.1523/JNEUROSCI.2604-19.2020>

Copyright © 2020 the authors

Uchida et al., 2012), suggesting a prominent role in reward-related behaviors since midbrain dopaminergic neurons convey reinforcement signals needed for learning (Schultz et al., 1997; Cohen et al., 2012). However, the precise contributions of distinct OT neuron types and subregions to projections to the VTA remains to be fully resolved (Heimer et al., 1987; Watabe-Uchida et al., 2012).

Physiologic evidence suggests that an important transformation in odor coding occurs in pPC and OT during sensorimotor decision-making. Activity of single neurons that correlate with choice and reward has been observed in both pPC and OT (Calu et al., 2007; Gadziola et al., 2015; Gadziola and Wesson, 2016). The anterior piriform cortex, which immediately precedes pPC and OT in the anatomic hierarchy of olfactory regions, does not have choice activity in the first sniff of an odor stimulus (Miura et al., 2012) when decision-making occurs (Uchida and Mainen, 2003; Abraham et al., 2004; Rinberg et al., 2006), although subsequent activity can reflect the decision (Gire et al., 2013). Single-neuron reward selectivity for single odors emerges before the motor response at the level of OT neurons (Gadziola et al., 2015); however, it is not known whether this is the case for pPC as well. If reward selectivity is computed in pPC during decision-making, OT could inherit this selectivity. Alternatively, heavy interconnection with the reward system provides the OT with the necessary reinforcement signals and plasticity-inducing dopamine for learning arbitrary odor–reward associations (Gadziola et al., 2015; Wieland et al., 2015). Indeed, responses of neurons to visual stimuli in dorsal striatum strongly reflect a newly learned reward association with a few trials of successful learning (Schultz et al., 2003; Pasupathy and Miller, 2005), so the olfactory striatum could function similarly to learn odor–outcome associations.

We developed a novel odor–reward categorization task to investigate the contributions of pPC and OT to odor-driven sensorimotor learning and decision-making. We recorded the activity of single neurons in pPC and OT as mice learned, through trial and error, the reward valence assigned to a panel of previously unexperienced odor stimuli. We also measured sniffing to focus on the earliest components of sensory processing and decision-making during the first sniff of odor cues. Mice reliably learned odor–reward contingencies within a single session, permitting the evolution of neural activity to be monitored throughout the learning process. We observed that reward selectivity was easily found in OT, multiplexed with odor selectivity, whereas explicit reward selectivity was largely absent in pPC. Together, these results support a striatal, rather than a cortical, model of rodent olfactory sensorimotor learning.

Materials and Methods

Animals (general). All experimental animals of either sex were C57BL/6 mice obtained from the Charles River Laboratories and were 2–4 months of age at the start of the experiments. Following surgical implantation of a tetrode drive and a custom head plate, all mice were rehoused individually. Experiments were conducted in accordance with Harvard University Animal Care Guidelines.

Surgery. Surgeries were performed on naive animals, and all behavioral training began after recovery from surgery. Mice were anesthetized with an intraperitoneal injection of a mixture of xylazine (10 mg/kg) and ketamine (80 mg/kg). To access sniffing information, airflow was measured through a cannula implanted in the nasal canal. A craniotomy (~1 mm) was made over the nasal canal (2 mm anterior of nasal/frontal fissure, 1 mm lateral) on the skull, with a goal of implanting a nasal cannula to monitor sniffing during behavioral and physiological experiments. An 18 gauge stainless steel cannula (~5 mm in length) was

inserted into the craniotomy. To affix the cannula to the skull, superglue was used for initial placement followed by two applications of dental cement to ensure stability of the cannula. Additionally, a craniotomy (~1 mm) was made over the dorsal skull at a location directly above the areas targeted for electrophysiological recordings with the goal of implanting a tetrode bundle (six tetrodes plus a 200- μ m-diameter optic fiber to ensure stability). The target locations were as follows: pPC (coordinates: 0.5 mm posterior and 3.8 mm lateral from bregma, and 3.8 mm ventral from brain surface) and OT (coordinates: 1.2 mm anterior and 1.5–2 mm lateral from bregma, and 4.6 mm ventral from brain surface). To ensure stability of the head of the animal during behavior and recording at a later stage, a custom-made head plate (made of light-weight titanium; dimensions, 30 $\times$ 10 $\times$ 1 mm; weight, 0.8 g) was affixed to the skull. A shallow well was drilled over the posterior lateral skull, and a single skull screw was affixed at that location. A wire was attached to this skull for grounding electrophysiological recordings. At the end of the surgery, a removable plastic cap was placed on the top of the nasal cannula to prevent foreign objects from entering the cannula. In addition, a plastic cone was positioned around the tetrode drive, and capped with a removable lid, to prevent damage to the drive. The nasal cannula and tetrode drive were implanted during the same surgery, but always on opposite (i.e., contralateral) sides. Each lateral olfactory tract delivers sensory input from a single olfactory bulb to the ipsilateral pPC and OT. Implanting the nasal cannula and tetrode drive on opposite sides was intended to avoid altering the natural airflow to the relevant sensory neurons for our recordings. Following the completion of the surgery, mice were given 1 week to recover.

Odor stimulus delivery. A custom-made 16-odor olfactometer was used to deliver stimuli during behavioral tasks. The olfactometer had a carrier stream of air calibrated to flow at 1 L/min. A single odor at a time could be added onto this carrier stream by opening an odor-specific valve to permit airflow (at 0.1 L/min) from an input manifold, through a tube containing liquid-phase odorant, and finally into the carrier stream. One-way check valves prevented the flow of odor from odorant-containing tubes into the carrier stream or back into the input manifold when air was not actively being flowed. Odor delivery to the animal was gated by a single (“final”) valve that directed odor to the animal or, between odor presentations, to an exhaust system. Note that this single valve was common to all odors and thus was not informative about odor identity. When odor was not flowed to the animal, a stream of clean air of the same flow rate was directed to the animal. This ensured active clearing of odors between trials. In addition, an exhaust system cleared odors from the behavior chamber. To ensure immediate delivery of odors, and comparable odor concentration from trial to trial, the line to the final valve was primed with odor from the current trial beginning at the end of the previous trial (i.e., minimum of 5 s, sufficient to replace the line volume three to five times). Sound from the switching of an odor-specific valve to perform selection of an odor for the upcoming trial was masked by a substantially louder non-odor-specific valve turned on simultaneously. The valves corresponding to each odor were randomly switched occasionally between sessions to prevent animals from learning sound cues, rather than odor cues. Finally, all tubing that was odor specific was replaced at the same time as new odors were added to the olfactometer to prevent odor from previous sessions lingering and being used as cues for task performance.

Behavioral task. When animals fully recovered from surgery (~7 d), they were placed on a water deprivation schedule for behavioral conditioning. Animals were first acclimated to head restraint (with their implanted head plate held in place) on the behavior rig over the course of 3–5 d. Then, animals were trained to lick for water from a spout positioned 2–3 mm from the mouth until satiated (typically, 1–1.5 ml); licking this spout (or not licking) would subsequently serve as the response for all behavioral tasks (Go/No-Go). Lick training took ~2–4 additional days. Next, odors were presented for 1 s with a varying (3 s minimum plus an exponential distribution with mean of 5 additional seconds) interstimulus interval, and reward was only available for a subset (half) of these odors. At first, only two odors (one rewarded, one unrewarded) were presented to expedite learning. The number of odors was increased

Figure 1. Behavioral task design and mouse performance. **A**, Top, Mouse in the head-fixed behavior setup. The odor port (red dot) and lick port (yellow dot) are labeled. Bottom, Diagram of the task structure, indicating the periods for odor, response (lick/no-lick), and reward delivery. S+ and S− denote rewarded and unrewarded stimuli respectively. **B**, Average task performance over the first four sessions for 14 mice. The hit rate, correct rejection (CR) rate, and overall performance rate are shown separately. **C**, Raster plots of individual lick times by trial, shown separately for rewarded and unrewarded odors for an expert mouse. **D**, Lick rate by trial type averaged over sessions in well trained mice (i.e., after session 5). Shading indicates SEM. FA, False alarm; CR, correct rejection. For **C** and **D**, 0 marks onset of odor stimulus. **E**, Top, An example trace of 5 s of respiration measurements obtained through air pressure from a nasal cannula. Exhalation is positive-going pressure, and inhalation is negative-going pressure. Bottom, Average sniff rate over all trials from an example behavior session. **F**, Histograms of the times of the first, second, third, and fourth sniffs of each trial.

to four and then eight (i.e., full task odor panel). Initially, water reward would be available immediately on licking during odor presentation. Gradually, the delay between lick/odor presentation and reward delivery was increased until reward delivery occurred at least 500 ms after the conclusion of odor presentation. To discourage a high false alarm rate, a hit was scored only on trials that the mouse licked in at least three of four 350 ms bins, beginning at the time of odor onset. It follows that mice were required to make a decision before the end of the second bin, 700 ms into the odor presentation, but this did not seem to force early inaccurate decisions.

Neural recordings began once the animals reach strong performance (>90% correct) on the full eight-odor panel. No odors were replaced during this initial period to facilitate the acquisition of task structure. In the final stage of experiments, a subset of odors was replaced between some sessions to study the learning of novel odor–reward associations.

Respiration monitoring. Respiration was recorded through a nasal cannula implanted during surgery. Following surgery, the nasal cannula was cleaned daily to prevent clogging. In this way, nasal cannulae typically stayed clear continuously for months of recordings. During experiments, a plastic tube connected to a pressure sensor was fitted over the nasal cannula to form a continuous pressure environment. Pressure signals were amplified 10×, bandpass filtered 0.1–100 Hz, and recorded at 1000 samples per second. For analysis, the start of each inhalation was identified as a negative-going zero crossing in the pressure signal. Likewise, the end of each inhalation (i.e., start of exhalation) was identified as a subsequent positive-going zero crossing in the pressure signal.

Electrophysiology. Neural activity was recorded with drives containing six tetrodes (Gray et al., 1995). The tip of each wire was gold plated until the impedance was 250–450 kΩ. Electrophysiological signals were acquired and amplified with a custom-made system built on two 16-channel analog chips designed by Intan Technologies. Each channel was digitized, sampled at 20 kHz, and bandpass filtered with a second-order Butterworth filter at 500–3000 Hz. Potential spike events were detected as any activity that crossed 3.7 SDs of noise (corresponding to 1:10,000 events by chance). Potential spike events were then manually clustered with MClust software (David Redish) to identify single-unit activity. The quality of clustering was ensured by requiring <1 in 1000 spikes to have occurred within a 2 ms refractory period of one another, clear overlap of single-spike waveforms, and L-distance <0.05. To minimize selection bias, all spikes that were visible during each session were recorded, sorted, and analyzed. At the end of the recording session each day, the entire bundle of tetrodes was lowered 40 μm to obtain a new set of neurons for the subsequent day. Following the completion of all behavior and recordings, tetrode placement was confirmed with an electrolytic lesion and postmortem histology. Reconstruction was attempted on all mice, but the electrolytic lesion location was only identifiable in five of six pPC mice and five of eight OT mice. We note that tetrodes were implanted with stereotactic technique to the same coordinates for each brain region across all animals, and we anticipate the remaining locations to also be on target (although we cannot be sure).

Area under the receiver-operating characteristic curve analysis. To character-

ize the responses of neurons to odors, we used the metric of area under the receiver-operating characteristic curve (auROC) for each cell–odor pair relative to baseline (i.e., prestimulus) firing rate. The auROC gives a measure of how well the firing rate at any given time can be discriminated from the baseline firing rate for that cell, independent of the absolute firing rates. Briefly, the value of the auROC for a given time bin indicates the percentile of that time bin in the distribution of baseline firing rates for bins of the same width. Therefore, a firing rate that is exactly the median of the baseline firing rate distribution would have an auROC of 0.5. Excitatory responses (yellow) correspond to an auROC >0.5 but ≤1, whereas inhibitory responses (blue) correspond to an auROC of <0.5 but ≥0. For analysis as well as data visualization, the auROC provides a more consistent measure to compare across neurons than absolute firing rate.

Decoding. Decoding analysis (both identity and reward category) was done with Support Vector Machines (SVMs) with a linear kernel, using the standard 80% training, 10% validation, and 10% test. Random shuffles from the entire population of neurons in a given region allowed many combinations, and therefore error estimates for decoding accuracy. Responses were obtained in a standard window of 200 ms from the onset of odor and first sniff inhalation (but see Fig. 4A, where this window was systematically varied). Shuffle controls for decoding reward category were obtained by scrambling the valence assignment across the odors.

Experimental design and statistical analysis. Nonparametric Wilcoxon signed-rank tests were calculated using the signrank function in MATLAB. Data are available on request.

Results

Mice quickly learn to categorize rewarded and unrewarded odors

We trained head-restrained mice to decide, based on the identity of an odor stimulus, whether to respond (Go) or not respond (No-Go) by licking a water port (Fig. 1A; see Materials and Methods). A panel of eight odors (four rewarded and four unrewarded) was presented on randomly interleaved trials during each experimental session. This panel size enabled us to characterize the sparseness of odor tuning while yielding 20–30 repetitions per odor to accurately measure behavioral and neural responses to the odor.

Mice acquired the task structure and reached a high level of performance (>90% correct trials) within three to four sessions from the start of training with odors (one session per day, consecutive days; Fig. 1B–D; 14 mice). Among correct behavioral responses, the fraction of Go responses on rewarded odor trials (i.e., hit rate) started higher and peaked earlier than the fraction of No-Go responses on unrewarded odor trials (i.e., correct rejection rate), reflecting an initial behavioral bias toward licking. Upon further training, mice consistently began licking within 200–400 ms of odor onset (Fig. 1C) for rewarded odors and maintained licking until reward delivery but did not even initiate licking on most unrewarded odor trials (Fig. 1C), indicating that the animals became confident of the reward association early during odor presentation. In a typical session, mice worked for ~300–500 trials and collected 150–250 water rewards before they were satiated.

Naive mice were initially trained on the task with the same panel of eight odors each day (one session per day) until they achieved successful performance (>90% correct) for two consecutive sessions. Following successful learning of the task structure, we replaced a subset of the odors with novel odors to study learning of odor–reward associations, which were typically learned to near-saturating (>90% correct) performance within one or two subsequent sessions. This process was repeated in daily sessions that continued for 1–2 months in the same mouse. Thus, the rapid learning of novel odor–reward associations permitted the observation of behavior and physiology during the acquisition of many odor–reward pairs. We first present analysis from sessions with odor–reward pairs in which successful learning had already been demonstrated through saturated high behavioral performance in a previous session (“familiar odors”) to characterize odor and reward tuning properties in

Figure 2. Putative recording sites and single-unit firing properties in pPC and OT. **A**, Recording sites estimated from tetrode tracks and electrolytic lesions in histologic sections. **B**, **C**, Baseline firing rates for all cells recorded in pPC and OT, respectively. Note: the last bin in **C** shows cells that had a firing rate of ≥ 60 spikes/s. **D**, **E**, Spike rasters aligned to odor onset (top) and aligned to the first sniff after odor onset (middle) as well as the PSTHs (bottom) with each trial aligned to the first sniff for an exemplar cell from pPC (**D**) and OT (**E**). Four PSTHs for the rewarded odors are shown in red, and four PSTHs for the unrewarded odors are shown in black.

pPC and OT. Subsequently, we describe the changes in behavioral and neuronal responses during the sessions in which novel odor–reward pairs were introduced and learning took place (“novel odors”).

Since neural responses to odors are aligned to respiration (Cury and Uchida, 2010; Shusterman et al., 2011), and behaving rodents can alter the rate of sniffing significantly (Welker, 1964; Kepecs et al., 2007; Wesson et al., 2008), we monitored respiration with a cannula implanted over the nasal cavity, contralateral to the side where tetrode recordings were performed (Fig. 1E). Once familiar with the task structure, mice timed their sniff to the onset of the odor cue, but the average number of sniffs in the 1 s odor period was unaltered (3.8 ± 0.52 vs 3.9 ± 0.45 s). The distribution of sniff times during odor presentation are shown in Figure 1F.

Figure 3. Responses of pPC and OT neurons. **A**, Coefficient of variation of neuronal responses in pPC and OT, separated for rewarded and unrewarded odors. For each panel, the median value is indicated by a dashed red line and the number next to it. **B**, Lifetime sparseness for cells recorded during sessions in which the animal was familiar with all eight odors in the stimulus panel. Sparser (more selective) cells have values closer to 1. **C**, **D**, auROC time course for each cell–odor pair for cells recorded in pPC and OT separated into rewarded and unrewarded odors. Odor onset is at 0 s. Yellow indicates that the cell was excited relative to its baseline firing rate, whereas blue indicates that the cell was inhibited relative to its baseline firing rate. For visualization purposes, cell–odor pairs with noisy firing during preodor time are not shown, resulting in a different number of cell–odor pairs between rewarded and unrewarded odors despite equal number of recorded cell–odor pairs. **E**, Population-averaged PSTHs for all recorded cell–odor pairs in pPC, split by rewarded and unrewarded odors. Error bars are SEM. **F**, Similar plot as in **E**, for OT neurons.

Task-related activity in pPC and OT

We used multitetrode drives (Gray et al., 1995) to isolate and record single units in either pPC ($n=6$ mice; 385 isolated units) or OT ($n=8$ mice; 270 isolated units). Following the completion of all behavior and recordings, tetrode placement was confirmed with an electrolytic lesion and postmortem histology (Fig. 2A). Baseline firing rates were different in the two areas (Fig. 2B,C), with pPC neurons having significantly lower firing rates than OT neurons (pPC: median = 0.64 Hz; range = 0–43.33 Hz; $n=372$ units; OT: median = 9.02 Hz; range = 0–107.02 Hz; $n=270$ units; rank sum = 85,888; $p=2.12 \times 10^{-64}$, Wilcoxon rank sum test).

Peristimulus time histograms (PSTHs) are shown for representative pPC and OT cells in Figure 2, D and E. Since inhalation of odors is necessary to evoke sensory responses, all of the remaining analyses were performed using the time of the first sniff as the start time for each trial. Example spike rasters show neuron activity before and after aligning to the first sniff. Example odor responses from a neuron in pPC (Fig. 2D)

highlights the lower baseline firing rate and sparser responses. More sustained responses can also be found (Fig. 3C,D).

Trial-to-trial variability in firing rate changes in response to odors was different in the two brain regions. The coefficient of variation was generally higher in pPC (rewarded: median = 2.01; range = 0.54–10.76; unrewarded: median = 2.14; range = 0.45–31.61) than in OT (rewarded: median = 1.11; range = 0.26–13.50; unrewarded: median = 1.26; range = 0.32–9.09; Fig. 3A). We characterized odor selectivity in the recorded neurons using the lifetime sparseness measure (Willmore and Tolhurst, 2001), which does not require thresholding or other selection criteria for determining response significance. Neurons in both regions spanned a wide range of selectivity, but neurons in OT had a lower selectivity than those in pPC (pPC: median = 0.45; range = 0.07–0.99; OT: median = 0.24; range = 0.00–0.67; rank sum = 3570; $p=2.65 \times 10^{-7}$, Wilcoxon rank sum test; Fig. 3B). These data indicate that pPC neurons have lower baseline firing rates, and more variable and more selective responses to odors than neurons in OT.

Reward-related activity at the single-neuron level

To characterize the responses of neurons to odors more objectively, we used the metric of auROC for each cell–odor pair relative to baseline (i.e., prestimulus) firing rate (Green and Swets, 1966; see Materials and Methods). We show the responses for all significant cell–odor pairs (at least one 100 ms bin different from baseline at a significance level of $p < 0.01$) in both areas with rewarded and unrewarded odors plotted separately (Fig. 3C,D). With this criterion, 69 units in pPC and 177 units in OT responded to at least one odor. This significance measure was used only for visualization and initial characterization, and further analysis (e.g., decoding) used all recorded neurons.

Odor responses in pPC were predominantly excitatory (Fig. 3C). Many cells continued to respond for the duration of the 1 s odor presentation, seen as a sustained band of yellow (Fig. 3C). Other cells had a strong transient (100–200 ms) response well aligned to the first inhalation of an odor, but did not respond for the remainder of the odor presentation despite continued sniffing (Fig. 3C). Interestingly, the cells that were inhibited appear to be inhibited by all odors nonselectively, corroborating previous findings (Otazu et al., 2015; Bolding and Franks, 2017). Low baseline firing rates made it difficult to detect odor-selective inhibition, if it were present, and suggest that these neurons did not directly contribute to odor discrimination.

A substantial portion of OT neurons were inhibited during odor presentation, often following a brief period of excitation (Fig. 3D). Transient excitatory responses at the beginning of odor presentation were much more prevalent in the OT (Fig. 3D).

We characterized reward coding in the pPC and OT, first focusing on familiar odors (i.e., odors that the mouse successfully learned in a previous session). Responses lasting >500 ms were more common in OT for rewarded odors than unrewarded odors (Fig. 3D). In contrast, the overall distribution of response profiles did not differ between rewarded and unrewarded odors for the population-averaged response across all pPC neurons (Fig. 3E). This difference is visible in the global average responses for rewarded versus unrewarded odors for both areas, with the population-averaged response of OT neurons (but not pPC neurons) showing divergent responses starting at ~ 300 ms after onset of the first sniff (Fig. 3E). This late divergence between average rewarded and unrewarded auROCs for OT cannot account for the decision of the mouse, which is made sooner (see below). The overlap of population-averaged PSTHs to rewarded and unrewarded odors in both pPC and OT suggests that a naive readout of average spiking activity from either population is an ineffective strategy for decoding reward value.

We then examined the extent to which reward information was explicitly available in the firing rate of individual neurons. Firing rates of individual neurons in each area suggested that individual OT neurons may be more discriminating between rewarded and unrewarded odors than pPC neurons (Fig. 4A,B). An auROC analysis comparing responses to all rewarded odors versus responses to all unrewarded odors revealed that reward selectivity was substantially more prevalent in OT than pPC both during and after odor presentation (Fig. 4C,D). Furthermore, approximately equal numbers of OT cells responded with higher firing rates for rewarded odors than for unrewarded odors (Fig. 4D, yellow) versus higher firing rates for unrewarded odors than rewarded odors (Fig. 4D, blue).

We then asked whether reward selectivity emerged at times corresponding to particular task events, such as stimulus onset

Figure 4. Reward valence selectivity during decision-making. **A**, Responses of three exemplar cells from pPC to eight odors, with no clear reward category information. **B**, Three exemplar cells from OT, two of which (red and blue) show clear reward selectivity (visualized by dashed lines separating firing rates), and one (black) showing imperfect reward selectivity. **C**, **D**, auROC to illustrate the ability of cells to discriminate rewarded versus unrewarded odors that had been successfully learned in previous sessions (i.e., familiar odors). Responses are aligned to the first inhalation following odor onset ($t = 0$). The cells (rows) are hierarchically clustered by the first three principal components calculated from the initial (first 200 ms) firing rate, the delayed response rate (200–800 ms), and the response rate at the time of reward delivery (1500–2000 ms). **E**, Histograms of the first time bin in which the auROC for a cell was significant for cells in pPC and OT. **F**, Zoom-in of the histograms in **E**, along with an overlay (hollow black bars) of the mean time of the first lick on hit trials to illustrate the beginning of the motor response.

or reward delivery. In order to measure the significance of reward selectivity, we compared the auROC at times after odor onset against the distribution of auROCs during prestimulus time. Consistent with a lack of actual reward selectivity for pPC cells, the times at which the reward auROC first reached a very liberal threshold of significance ($p < 0.01$, 35 comparisons/cell for the 100 ms time bins spanning the 3.5 s window; familywise type 1 error rate = 0.296) were scattered throughout the trial and were not well aligned to any task events (Fig. 4E, orange). In contrast, reward selectivity emerged in 46% of OT cells within the first 300 ms of odor sampling (Fig. 4E, blue). The absence of any subsequent peaks indicates that there was not a separate population of OT neurons that responded only to later task events, such as reward delivery. The majority of these selective responses emerged before the mouse began to lick on rewarded odor trials (Fig. 4F). In order to focus on pre-decision neural activity, all of the following analyses use a 200 ms response window beginning at the start of the first inhalation following odor onset.

Figure 5. Multiplexing of odor and reward codes. **A**, Single-cell peristimulus time histograms are used to produce barcodes that illustrate both odor and reward sensitivity of the response of a cell during the first 200 ms of odor presentation, beginning with the first inhalation. Odors are rank ordered from the most positive-going response to the most negative-going response, then colored according to valence (red, rewarded; black, unrewarded) and the normalized response to the odor. **B**, Stacks of barcodes for cells recorded in pPC and OT during sessions in which all eight odors in the stimulus panel had been successfully learned in a previous session (i.e., familiar odors). **C**, Mean reward selectivity, as measured by auROC, as a function of lifetime sparseness. auROC was calculated for the first 200 ms from the start of odor sampling (first inhalation following odor onset). Error bars are SEM. The number of cells used to calculate each point in the plots are indicated in parentheses. **D**, The percentile of the true valence auROC (from a nondiscriminating value of 0.5) among the distribution of auROCs generated from shuffling the valence labels of the odors. Dots are individual cells, the connected lines show medians, and the error bars are 90% confidence intervals of the median. The dashed line indicates chance; that is, the true auROC falling at the median of the distribution on average. The number of cells used to calculate each point in the plots is indicated in parentheses.

Multiplexing of odor and reward codes in OT at the single-neuron level

To visualize the relationship between odor and reward selectivity in single neurons, we created a bar code for each neuron that separates responses to each of the eight individual odors but also distinguishes odors based on their reward valence (Fig. 5A). The position of each odor in the bar code corresponds to the rank order of the trial-averaged response during the first 200 ms window, from most positive on the left to most negative on the right. The color of each segment corresponds to the valence of the odor (red, rewarded; black, unrewarded) and the intensity of the

color corresponds to the odor-evoked response magnitude (i.e., absolute value of spike rate during odor sampling minus baseline) normalized per neuron to the strongest response among the eight odors. The combined barcodes in pPC illustrates that neurons have sparse selectivity for odors that limit the ability of individual neurons to discriminate valence. In OT, neurons respond much more densely to odors with the trial-averaged odor responses of 30 of 81 neurons perfectly discriminating valence (far exceeding chance under a binomial model equal to 1 perfect discriminator of every 35 neurons) and the majority of OT neurons partially, but not entirely, discriminating valence (Fig. 5B).

The auROC metric for reward selectivity is susceptible to overestimating reward selectivity for cells that respond to only a few odors that happen to be assigned the same reward valence by chance. Indeed, neurons with greater odor selectivity did tend to show greater reward selectivity as measured by auROC, and pPC had more of these odor selective neurons than OT leading to higher apparent reward selectivity in pPC (Fig. 5C). To control for odor selectivity, we compared the observed reward selectivity of each cell to a null distribution of 10,000 auROCs generated by sampling, with replacement, responses to eight odors and randomly assigning four rewarded labels and four unrewarded labels regardless of the actual experimental reward value. This control revealed that pPC cells did not have reward selectivity above chance, whereas OT cells did despite lower overall auROC values (Fig. 5D; pPC, 4.74%, OT, 48.18%; at $p < 0.05$ significance threshold). Therefore, reward selectivity was detectable at the single-neuron level in OT but not pPC.

We then performed a complementary analysis to test whether single neurons in pPC or OT exhibited odor selectivity independent of reward selectivity. In this case, we calculated eight auROCs for each neuron, each auROC comparing one odor against the other three with the same reward valence. We generated two null distributions of 10,000 auROCs for each neuron, one for rewarded odors and one for unrewarded odors, by sampling, with replacement, from the pool of trial responses to odors of the same reward valence; only trials with correct behavioral reports were included to control for behavioral variation across odors. Response selectivity for at least one of eight odors was detectable in a substantial portion of pPC and OT neurons (pPC, 44.74%; OT, 46.36%; at $p < 0.05$ significance threshold with Bonferroni correction for eight comparisons).

Together, these statistically validated selectivity metrics allowed us to determine whether odor identity and reward valence were conveyed by overlapping or distinct sets of OT neurons. The fraction of OT neurons exhibiting both reward and odor selectivity was substantial and indistinguishable from the expectation for statistically independent (not mutually exclusive) reward-selective and odor-selective neuronal populations (OT, 22.73% selective for both odor and reward; vs 22.34% expected for statistical independence; pPC, 0.00% selective for both odor and reward vs 2.12% expected for statistical independence). This analysis indicated that single neurons in OT carry information about odor identity multiplexed with reward information.

Reward coding at the population level

Collectively, many individual cells with weak reward valence selectivity could still give rise to reliable valence selectivity at the population level. To determine whether our pPC and OT datasets had this property, we attempted to decode the valence of odors from single-trial responses of cells in each area using a support vector machine with a linear kernel (Hastie et al., 2009). For each cell, the trials during the session it was recorded were randomly assigned to training (80%), validation (10%), and test (10%) sets.

To empirically determine an appropriate time window to analyze, we decoded reward valence from OT activity for a series of windows beginning at the first inhalation of the stimulus; specifically, we used the average firing rate response (i.e., baseline subtracted) for each cell within the window. Decoding performance improved with progressively longer time windows until saturation at ~200 ms (Fig. 6A). Accordingly, all subsequent decoding analyses were performed using the average firing rate response within a 200 ms window beginning at the first sniff of the odor.

To directly test for an explicit population code for reward valence, we trained linear classifiers (SVMs) to decode reward from the responses to a set of four odors (two rewarded, two unrewarded), then tested the ability of the same classifier to decode reward from the responses to the other set of four odors presented in the same session (Fig. 6B). Overall, we found that classifiers trained on OT data successfully decoded reward from responses to odors that the classifiers did not see during training. On the other hand, classifiers trained on pPC data did not have this property. Importantly, these results were consistent across individual mice (OT, $n = 5$ mice; pPC, $n = 3$ mice) and odor sets for which populations of at least five cells were available (Fig. 6B; Materials and Methods).

A single neuron could respond similarly to two odors with the same valence, because the response is valence specific or strictly due to similar purely sensory-driven responses evoked by the two odors. In turn, the activity of a population of neurons might also falsely appear to discriminate rewarded versus unrewarded odors due to chance (i.e., the presence of sensory but not reward information) when probing with a small number of odors. If population activity explicitly codes for reward valence beyond sensory tuning alone, however, a single classifier should be able to decode the reward valence for increasingly larger number of odors without performance declining because the reward information persists despite the presence of extraneous stimulus-specific information. Therefore, we tested the ability of a decoder to classify reward valence from a progressively larger number of odors (e.g., one rewarded odor vs one unrewarded odor, two rewarded odors vs two unrewarded odors). To minimize the contribution of stimulus tuning for any single odor to the overall ability to successfully categorize rewarded versus unrewarded

Figure 6. Decoding reward valence from population data. **A**, Decoding accuracy for reward valence obtained with a support vector machine, trained and tested on the average response rate of OT cells in a window that began at the time of the first inhalation of an odor and had a duration equal to the time denoted for each curve. **B**, Training and testing linear classifiers on different sets of odors. Support vector machines with a linear kernel were trained on the single-trial responses to one set of four odors and tested on a different set of four odors. Each line corresponds to the test performance for a different mouse or set of test odors. pPC (yellow): three curves total, each curve is a different mouse (each mouse had a different set of test odors); OT (blue): five curves total, with three curves for three mice with the same test odor set, one curve for a second set of test odors for one of the three mice, and one curve for a fourth mouse with a different set of odors than the other three mice. **C**, **D**, Decoding accuracy for similar classifiers that were trained and tested on the given number of odors (color correspondence indicated in the inset), half of which were rewarded odors and half of which were unrewarded odors. **E**, **F**, The same as **C** and **D** but for classifiers trained and tested on odor responses with the reward valence label shuffled. **G**, Decoding of reward in OT and pPC for two, four, six, and eight odors (same color scheme as in **C** and **D**) among cells with lifetime sparseness between 0.3 and 0.6. This was the range of lifetime sparseness in which pPC and OT most overlapped (Fig. 5A). OT still performed better than pPC even when lifetime sparseness in the two areas was matched in this way. **H**, Comparison of the reward-decoding performance for eight odors in OT (blue) and pPC (yellow) for the same data as **G** but with the reward labels shuffled. All error bars are SEM performance across 100 classifiers, each of which was trained and tested on a random subsample of the dataset corresponding to the population size indicated on the horizontal axis.

odors, we leveraged the fact that the data collected from each mouse comprised many odor–reward associations, aggregated from individual sessions with eight odors each. We therefore trained and tested the classifiers on this pooled dataset with balanced numbers of rewarded and unrewarded odors (14 mice;

Figure 7. Learning of novel odor–reward associations. **A**, During the initial phase of training on the task structure, the eight odors in the stimulus panel remained the same. After 5–10 sessions, half of the odors (two rewarded and two unrewarded) were replaced between some sessions (“Novel”). Before changing the stimulus panel again, the stimulus panel was held constant for at least one additional session to measure behavioral and neural responses to the newly familiar odors. Different shapes indicate different sets of odors, with odors represented by the same shape being introduced at the same time. Hollow shapes indicate novel odors that the mouse has not been exposed to in previous sessions. Filled shapes indicate odors that have been learned in a previous session (“Familiar odors”). Red shapes indicate rewarded odors, and black shapes indicate unrewarded odors. **B**, The performance for the four novel odors on the first 3 d of learning is shown averaged for all novel odors as a function of presentation number. Each novel odor was presented ~35 times (randomly) in each session. **C**, **D**, Go/No-Go behavioral responses of mice to novel odors in the first session: red for lick and black for no-lick. Each row is one novel odor. The length of each row corresponds to the total number of presentations of that particular odor during its first session in the odor panel.

minimum of five sessions per mouse). Importantly, the size of the training set (i.e., number of single-trial responses) was held constant regardless of the number of odors used.

In this test, the decoder successfully classified pPC responses to two odors (one rewarded and one unrewarded), but performance declined rapidly when challenged to decode valence from responses to larger numbers of odors, suggesting that pPC odor responses conveyed little to no linearly decodable information about valence (Fig. 6C). As a control, we compared this result to the decoding performance achieved by a classifier that was trained and tested on the same dataset after shuffling the valence labels of the odors. This shuffling procedure would remove any potential explicit reward information and force the decoder to rely on stimulus tuning alone for classification. Decoding performance on valence-shuffled pPC data followed the same trend as the unshuffled pPC data (Fig. 6E), again suggesting the absence of an explicit code for reward in pPC. In contrast, decoder performance for the OT data did not decrease substantially as the number of odors increased (Fig. 6D), but did decrease for valence-shuffled OT data, confirming that valence-specific information was lost in the shuffling process (Fig. 6F). We also limited the decoding analysis to cells with lifetime sparseness between 0.3 and 0.6, which was the range in which pPC and OT most overlapped (Fig. 5C). We then decoded reward for 2, 4, 6,

and 8 odors, and found that OT still performed better than pPC even when lifetime sparseness in the two areas was matched in this way (Fig. 6G). As controls, we compared reward decoding performance for eight odors in OT and pPC for the same data, but with the reward labels shuffled (Fig. 6H). While the shuffled data fare very similar to real data for pPC, they are much worse for OT, again confirming that OT neurons have explicit representation of reward category, even after normalizing for response sparsity.

These results clearly indicate that the reward category of odors could be explicitly read out from small groups of OT neurons, but not from pPC neurons.

Novel odor learning

Next, we investigated the evolution of reward coding during the process of learning novel odor–reward associations, which we were able to do within recording sessions. Approximately every other session following successful learning of the task structure (total of 71 sessions), four of the eight odors that had been used in previous sessions were replaced with four novel odors (two rewarded and two unrewarded; Fig. 7A). Mice were not exposed to the odors before training and reward valence was randomly assigned to each odor for every mouse.

Figure 8. Neural response changes during novel odor learning. **A**, Responses of an example neuron to the 4 rewarded (S+) and 4 unrewarded (S−) odors, half of each were familiar and novel. Trials were randomly presented, but are shown sorted for visualization. Within each category, trials are displayed in order of presentation in the session, with early trials at bottom (arrow at bottom right). The accuracy of performance for each of the odors is indicated at right. Note that accuracy for familiar odors is near 1 for all trials, but for novel odors accuracy starts low and approaches 1. **B**, Based on the average firing rate (in 200 ms after first sniff) for each neuron–odor pair, the difference in firing rate between the last 10 and the first 10 trials was calculated for novel rewarded and unrewarded odors. As a population, neuronal responses for rewarded odors become slightly ($p < 0.05$) stronger compared with unrewarded odors. Each of the recorded neurons was only recorded during a single session, and therefore had a different stimulus panel than neurons recorded during other sessions. **C**, Decoding reward selectivity from OT responses to novel odors. Inset, A classifier was trained to discriminate two familiar rewarded odors from two familiar unrewarded odors, then tested on its ability to predict reward valence from responses to two novel rewarded and two novel unrewarded odors. Decoding accuracy (blue; averaged across 71 novel sessions) steadily increased with repeated presentation of novel odors, and tracked performance of mice (red). Error bars show SEM.

We found that across different mice and different novel odors (eight mice, 71 sessions) rapid learning occurs in the very first session, with additional improvement in performance over the next two sessions (Fig. 7B). Correct Go responses to rewarded odors become frequent earlier than correct No-Go responses to unrewarded odors (Fig. 7C,D), consistent with a previously observed behavioral bias toward Go responses (Fig. 1B). High accuracy is maintained in these sessions for the already familiar odors.

We recorded the activity of OT neurons during novel odor learning, with four familiar and four unfamiliar odors presented in randomly interleaved trials (Fig. 8A). We asked whether the activity of individual neurons changed during the course of the session by examining the difference in firing rate for the first and last 10 trials of an odor. For each OT cell recorded during learning, we calculated the change in response to each novel odor normalized to the mean and SD of the cell responses to the four familiar odors used in the same session. Across the population of neurons, we found that the distribution of these normalized response changes for rewarded novel odors (mean change, 0.22 ± 0.13 spikes/s; $n = 69$ cells) was significantly different ($p = 0.02$, comparison of distributions with a Kolmogorov–Smirnov test) from the distribution for unrewarded novel odors (mean change, -0.22 ± 0.14 spikes/s; $n = 61$ cells; Fig. 8B).

To determine whether odor–reward associations were reflected in the activity of populations of OT neurons during learning, we leveraged the fact that only four of the odors were novel in learning sessions and the other four odors were familiar (i.e., learned in a previous session). We therefore compared the evolution of reward selectivity of novel odors using the familiar odors as a reference point (Fig. 8C, inset). To this end, we trained a linear decoder to classify single-trial responses (mean firing rate response in first 200 ms of odor sampling) to familiar rewarded versus unrewarded odors, then tested the performance on novel rewarded and unrewarded odors. First, we quantified the density of reward valence coding in OT by determining the minimum population size for which this training procedure resulted in decoding accuracy that approximated the average behavioral performance of a well trained mouse ($\sim 90\%$) on familiar odors alone; this size was 20 cells (Fig. 6D). The average performance of 100 classifiers each trained on a

random group of 20 cells of the 73 recorded during single-session learning is shown in Figure 8C. Indeed, there is a highly significant ($p < 10^{-6}$) increase in decoder performance as a function of the presentation number (i.e., trials with the same odor), starting near chance performance at the beginning of the session and reaching comparable performance to familiar odors ($>90\%$) by the end of the very first session with the novel odors. Remarkably, the time course of the improvement in decoding closely tracks the performance of mice in discovering the valence of the novel odors (Fig. 8C). Overall, these results establish that reward selectivity in the OT develops in parallel with learning at the behavioral level.

Relating neural activity and behavior

If the activity in OT is used to drive behavior, we would expect a stronger relationship between coding and behavior (i.e., “consistency”; Majaj et al., 2015) in OT than other candidate areas, such as pPC. Indeed, the relationship between the decoding accuracy for individual odors and behavioral performance of the mouse for the corresponding odors was closer to statistical significance in OT than pPC (in OT: $r = 0.40$; $p = 0.05$; in pPC: $r = 0.19$; $p > 0.2$; Fig. 9A,B).

Finally, we sought to rule out the possibility that OT responses were more closely related to the motor decision than the expected reward valence. The Go/No-Go task structure ties reward valence to motor decision (i.e., Go = rewarded, No-Go = unrewarded), so we examined OT responses on error trials to dissociate these variables. Because of the asymmetry in hit rate versus correct rejection rate in our task, there were too few misses to be used for this analysis; instead, we focused on false alarm trials. We quantified the reward valence coding by calculating the auROC of OT neuron responses across hit versus false alarm trials for which the motor decision was the same but the valence differed (“reward auROC”). Likewise, we quantified the motor decision coding by calculating the auROC of OT responses across correct rejection versus false alarm trials (“motor auROC”), for which the expected reward valence was the same but the motor decision differed. Importantly, we never included responses from the same odor among both the false alarm trials and correct rejection trials when calculating the

Figure 9. *A*, Odor-by-odor relationship between behavioral performance and neural coding for pPC. Each dot is a single odor. $N = 19$, $r = 0.19$, $p = 0.41$. *B*, Same plot as in *A*, for OT. $N = 24$, $r = 0.4$, $p = 0.05$. *C*, OT activity selectivity on error trials. The effect of motor (lick/no-lick) decision was measured as the auROC when discriminating responses during the first 200 ms of false alarm versus correct rejection trials. Similarly, the effect of reward valence was measured as the auROC when discriminating hit versus false alarm trials. A minimum of 15 false alarm trials was required for the cell to be included in the analysis. Dashed line is the unity line.

motor auROC. This ensured that we did not artificially decrease the motor auROC by forcing the discrimination of an odor from itself, relative to the reward auROC in which the odors on hit versus false alarm trials were always different, by construction. If motor decision was the main driver of selective responses in OT neurons, the motor auROC should be greater than the reward auROC, but we did not find this relationship (Fig. 9C; Wilcoxon rank sum test, $p = 0.269$; $n = 45$ neurons; ranksum = 2185). This result indicates that the observed reward tuning is not primarily driven by a difference in motor behavior.

Discussion

In this work, we developed an odor–reward categorization task in which mice could flexibly acquire novel odor–reward associations within a single session. We recorded neural activity in behaving animals and discovered that activity in pPC largely reflects sensory coding, with very little explicit information about categories. By contrast, the OT acquires a representation of reward category rapidly in a matter of minutes, and this information is expressed within 100 ms of odor sampling (well before motor action). Moreover, reward selectivity and successful behavioral performance developed in parallel as novel odor–reward associations were learned. These results provide important constraints on how odors are processed in the olfactory system during reward-motivated behaviors.

Identity coding

Many previous studies have focused on the coding of odor identity in the PC (Illig and Haberly, 2003; Stettler and Axel, 2009; Gottfried, 2010; Wilson and Sullivan, 2011; Bolding and Franks, 2017; Iurilli and Datta, 2017; Roland et al., 2017). By contrast, our experiments were designed to focus on reward-related activity, and only a small number of neurons were recorded during experimental sessions with a particular set of odors. Previous studies (Bolding and Franks, 2017; Iurilli and Datta, 2017) have indicated that a few dozen cells are necessary for odor identity decoding in the piriform cortex when 6–15 odors are considered. Identity decoding was very poor in pPC in our current study, likely due to the sparse odor responses in pPC resulting in few or no responses to a given odor within the relatively small populations of cells available for this analysis (mean, 32 cells per odor set). Individual OT neurons carry some information about both odor identity and reward category. This multiplexing of information in OT may allow rapid and flexible learning of reward-oriented behavior.

Reward coding

A major finding of our study is that neurons in the OT carry an explicit representation of reward category, independent of the identity of the odor. This information can be gleaned reliably from the firing rates of only a few dozen arbitrarily selected neurons, within a brief period (<200 ms) after odor sampling. Remarkably, when mice learn new odor–reward associations, neurons in the OT alter their firing properties to begin coding for reward category in concert with learning during the same session. Importantly, we show that the activity that is correlated with reward valence is not simply a consequence of the motor action. First, the emergence of reward information in the activity of OT neurons precedes the motor action (licking). Second, when using error trials, the ability to decode the reward category for each neuron is not dominated by the ability to decode the choice of the mouse (licking or not licking). Therefore, while OT neurons may also carry information related to the motor action, this would be in addition to the information they carry about reward category.

The OT is a heterogeneous structure, with different types of neurons and mesoscopic structures such as the islands of Calleja (Heimer et al., 1982; Wesson and Wilson, 2011; Giessel and Datta, 2014). Our experimental design did not allow us to distinguish among different cell types, but striatal cells dominate the OT in numbers (Heimer et al., 1982; Millhouse and Heimer, 1984). Additionally, cells in the islands of Calleja have thin and sparse dendrites, which might be unfavorable for generating strong dipoles to be easily detected by extracellular electrodes. Parts of the ventral pallidum can protrude into the OT, and we cannot be certain we avoided recording there. We note, however, that we did not find a relationship between the dorsal–ventral position of neurons in the tetrode path (tetrodes were lowered every recording session) and reward value selectivity (i.e., $|\text{auROC} - 0.5|$) in OT (Pearson correlation of depth vs valence selectivity: $r = -0.136$, $p = 0.072$; $n = 177$ neurons). Future work focusing on electrical or optical recordings from specifically labeled neurons can disentangle contributions from different neuron types and subregions.

There is significant anatomic and pharmacological work implicating the OT in reward and motivated behavior (Heimer et al., 1982; Ikemoto, 2003, 2007). Recent work has implicated the OT in valence coding (Gadziola et al., 2015; Gadziola and Wesson, 2016; Zhang et al., 2017a), though the relationship between valence coding and stimulus tuning was not clear. Here, we used a large number of odors to probe the breadth of stimulus responses when many different odors reliably predict reward or

no reward. We find that OT neurons multiplex odor identity and reward category information. Additionally, the behavioral performance of the mouse for an individual odor is correlated with the ability to decode reward category from OT activity in response to that odor. By monitoring sniffing, we could determine the timing of neural responses at high resolution and found that reward category-related activity in the OT arises rapidly (on average, 100 ms after the onset of sniffing, well before the average onset of the motor act). In contrast, we found little evidence for an explicit representation of reward category information in the pPC, beyond that which is expected merely from sensory tuning. This appears to be at odds with an earlier study (Calu et al., 2007) suggesting that the activity of pPC neurons can be modulated by reward. However, that study used only a few odors and only a few cells (less than five) showed the effect, leaving open the possibility that the responses they observed were due to sensory tuning alone (as demonstrated in Fig. 6) or reflected later-onset brain-wide reward consumption or anticipatory signals (Schultz, 2000).

How do OT neurons acquire reward-selective responses so early during learning?

The emergence of reward-selective responses in the striatum is not unprecedented. Monkeys trained to associate reward with visual stimuli develop reward selectivity in the dorsal striatum that is apparent in the activity of individual neurons within a few trials of successful learning (Schultz et al., 2003; Pasupathy and Miller, 2005). In contrast, neurons in the prefrontal cortex took much longer to acquire reward selectivity (Pasupathy and Miller, 2005). OT, at least in rodents, has been suggested to implement similar computations for olfactory information that the rest of striatum implements for other modalities (Wesson and Wilson, 2011; Giessel and Datta, 2014). Indeed, OT and dorsal striatum have a similar composition of cells (e.g., ~98% D₁ or D₂ receptor-expressing medium spiny neurons), comparable density of input from VTA, and the same output targets (Wesson and Wilson, 2011). Our results strengthen the analogy between OT and dorsal striatum, adding physiological data to the known anatomic and cytoarchitecture similarities.

Reward-selective responses might emerge in the striatum early during learning to support decision-making. Indeed, corticostriatal inputs have been shown to drive decision-making in an auditory discrimination task (Znamenskiy and Zador, 2013), and corticostriatal synapses exhibit plasticity over the course of training (Xiong et al., 2015). Moreover, the strength of olfactory synaptic inputs to the ventral striatum can be modulated by optogenetic release of dopamine in slices (Wieland et al., 2015), and the activation of VTA projections to the OT can induce odor preference in mice (Zhang et al., 2017a).

Sensory processing begins with a dedicated channel (glomerulus) for each olfactory receptor in the OB. An optimal linear readout of these channels is sufficient to determine the presence of any individual odor among at least a moderately sized (16 odor) panel of structurally similar odors and, thus, decide the appropriate motor action given a complex sensory scene (Mathis et al., 2016). The success of a linear readout implies that any monosynaptic recipient of OB output could learn to make accurate sensorimotor decisions through biologically plausible mechanisms such as reward-gated Hebbian plasticity (Reynolds and Wickens, 2002; Hung et al., 2005; Loewenstein and Seung, 2006; Pfeiffer et al., 2010). Therefore, one possible model for rapid learning of reward selectivity involves OB or PC inputs that evoke odor responses in OT being subsequently reinforced by

dopamine release from VTA projections to the OT on reward delivery.

A study was published recently during the preparation of this article that takes a complementary approach to examine the relationship between pPC and OT in coding odor–reward associations (Gadziola et al., 2019). Those authors use reward contingency reversals to demonstrate that odor-evoked responses in OT change in response to experimental manipulation of the reward valence predicted by the odor. Their results corroborate the much denser reward coding in OT than pPC observed in our work. In addition, our approach of training each mouse to learn many odor–reward associations enabled the dissociable characterization of odor and reward valence components of the tuning of the cells in the OT. Together, along with previous literature, our two studies lend strong support for a central role of the OT in odor-driven reward-motivated behaviors.

Our study sheds light on the representation of a behaviorally relevant aspect of the environment, reward, and how it relates to sensory coding in the olfactory system. The rapidity of odor-guided learning in mice and the direct connectivity of the OT to the sensory periphery, as well as motor and reward areas, make this behavior and brain region attractive candidates for the dissection of rapid sensorimotor reinforcement.

References

- Abraham NM, Spors H, Carleton A, Margrie TW, Kuner T, Schaefer AT (2004) Maintaining accuracy at the expense of speed: stimulus similarity defines odor discrimination time in mice. *Neuron* 44:865–876.
- Bolding KA, Franks KM (2017) Complementary codes for odor identity and intensity in olfactory cortex. *Elife* 6:e22630.
- Calu DJ, Roesch MR, Stalnaker TA, Schoenbaum G (2007) Associative encoding in posterior piriform cortex during odor discrimination and reversal learning. *Cereb Cortex* 17:1342–1349.
- Cohen JY, Haesler S, Vong L, Lowell BB, Uchida N (2012) Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* 482:85–88.
- Cury KM, Uchida N (2010) Robust odor coding via inhalation-coupled transient activity in the mammalian olfactory bulb. *Neuron* 68:570–585.
- Dayan P, Niv Y (2008) Reinforcement learning: the good, the bad and the ugly. *Curr Opin Neurobiol* 18:185–196.
- Diodato A, Ruinart de Brimont M, Yim YS, Derian N, Perrin S, Pouch J, Klatzmann D, Garel S, Choi GB, Fleischmann A (2016) Molecular signatures of neural connectivity in the olfactory cortex. *Nat Commun* 7:12238.
- Gadziola MA, Wesson DW (2016) The neural representation of goal-directed actions and outcomes in the ventral striatum's olfactory tubercle. *J Neurosci* 36:548–560.
- Gadziola MA, Tylicki KA, Christian DL, Wesson DW (2015) The olfactory tubercle encodes odor valence in behaving mice. *J Neurosci* 35:4515–4527.
- Gadziola MA, Stetzk LA, Milton AJ, Arakawa K, Wesson DW (2019) A neural system for encoding odor-reward associations which influences approach behaviors. *bioRxiv*. Advanced online publication. Retrieved April 21, 2020.
- Giessel AJ, Datta SR (2014) Olfactory maps, circuits and computations. *Curr Opin Neurobiol* 24:120–132.
- Gire DH, Whitesell JD, Doucette W, Restrepo D (2013) Information for decision-making and stimulus identification is multiplexed in sensory cortex. *Nat Neurosci* 16:991–993.
- Gottfried JA (2010) Central mechanisms of odour object perception. *Nat Rev Neurosci* 11:628–641.
- Gray CM, Maldonado PE, Wilson M, McNaughton B (1995) Tetrodes markedly improve the reliability and yield of multiple single-unit isolation from multi-unit recordings in cat striate cortex. *J Neurosci Methods* 63:43–54.
- Green DM, Swets JA (1966) Signal detection theory and psychophysics. New York: Wiley.

- Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning. New York: Springer.
- Heimer L, Switzer RD, Van Hoesen GW (1982) Ventral striatum and ventral pallidum. *Trends Neurosci* 5:83–87.
- Heimer L, Zaborszky L, Zahm DS, Alheid GF (1987) The ventral striatopallidothalamic projection: I. The striatopallidal link originating in the striatal parts of the olfactory tubercle. *J Comp Neurol* 255:571–591.
- Hung CP, Kreiman G, Poggio T, DiCarlo JJ (2005) Fast readout of object identity from macaque inferior temporal cortex. *Science* 310:863–866.
- Ikemoto S (2003) Involvement of the olfactory tubercle in cocaine reward: intracranial self-administration studies. *J Neurosci* 23:9305–9311.
- Ikemoto S (2007) Dopamine reward circuitry: two projection systems from the ventral midbrain to the nucleus accumbens–olfactory tubercle complex. *Brain Res Rev* 56:27–78.
- Illig KR, Haberly LB (2003) Odor-evoked activity is spatially distributed in piriform cortex. *J Comp Neurol* 457:361–373.
- Iurilli G, Datta SR (2017) Population coding in an innately relevant olfactory area. *Neuron* 93:1180–1197.e7.
- Johnson DM, Illig KR, Behan M, Haberly LB (2000) New features of connectivity in piriform cortex visualized by intracellular injection of pyramidal cells suggest that “primary” olfactory cortex functions like “association” cortex in other sensory systems. *J Neurosci* 20:6974–6982.
- Kepecs A, Uchida N, Mainen ZF (2007) Rapid and precise control of sniffing during olfactory discrimination in rats. *J Neurophysiol* 98:205–213.
- Loewenstein Y, Seung HS (2006) Operant matching is a generic outcome of synaptic plasticity. *Proc Natl Acad Sci U S A* 103:15224–15229.
- Majaj NJ, Hong H, Solomon EA, DiCarlo JJ (2015) Simple learned weighted sums of inferior temporal neuronal firing rates accurately predict human core object recognition performance. *J Neurosci* 35:13402–13418.
- Mathis A, Rokni D, Kapoor V, Bethge M, Murthy VN (2016) Reading out olfactory receptors: feedforward circuits detect odors in mixtures without demixing. *Neuron* 91:1110–1123.
- Millhouse OE, Heimer L (1984) Cell configurations in the olfactory tubercle of the rat. *J Comp Neurol* 228:571–597.
- Miura K, Mainen ZF, Uchida N (2012) Odor representations in olfactory cortex: distributed rate coding and decorrelated population activity. *Neuron* 74:1087–1098.
- Otazu GH, Chae H, Davis MB, Albeanu DF (2015) Cortical feedback decorrelates olfactory bulb output in awake mice. *Neuron* 86:1461–1477.
- Pasupathy A, Miller EK (2005) Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature* 433:873–876.
- Pfeiffer M, Nessler B, Douglas RJ, Maass W (2010) Reward-modulated Hebbian learning of decision making. *Neural Comput* 22:1399–1444.
- Reynolds JNJ, Wickens JR (2002) Dopamine-dependent plasticity of corticostriatal synapses. *Neural Netw* 15:507–521.
- Rinberg D, Koulakov A, Gelperin A (2006) Speed-accuracy tradeoff in olfaction. *Neuron* 51:351–358.
- Rokni D, Hemmelder V, Kapoor V, Murthy VN (2014) An olfactory cocktail party: figure-ground segregation of odorants in rodents. *Nat Neurosci* 17:1225–1232.
- Roland B, Deneux T, Franks KM, Bathellier B, Fleischmann A (2017) Odor identity coding by distributed ensembles of neurons in the mouse olfactory cortex. *Elife* 6:e26337.
- Schultz W (2000) Multiple reward signals in the brain. *Nat Rev Neurosci* 1:199–207.
- Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. *Science* 275:1593–1599.
- Schultz W, Tremblay L, Hollerman JR (2003) Changes in behavior-related neuronal activity in the striatum during learning. *Trends Neurosci* 26:321–328.
- Shusterman R, Smear MC, Koulakov A, Rinberg D (2011) Precise olfactory responses tile the sniff cycle. *Nat Neurosci* 14:1039–1044.
- Slotnick B (2001) Animal cognition and the rat olfactory system. *Trends Cogn Sci* 5:216–222.
- Slotnick B, Hanford L, Hodos W (2000) Can rats acquire an olfactory learning set? *J Exp Psychol Anim Behav Process* 26:399–415.
- Stettler DD, Axel R (2009) Representations of odor in the piriform cortex. *Neuron* 63:854–864.
- Sutton RS, Barto AG (1998) Reinforcement learning: an introduction. Cambridge, MA: MIT.
- Uchida N, Mainen ZF (2003) Speed and accuracy of olfactory discrimination in the rat. *Nat Neurosci* 6:1224–1229.
- Watabe-Uchida M, Zhu L, Ogawa SK, Vamanrao A, Uchida N (2012) Whole-brain mapping of direct inputs to midbrain dopamine neurons. *Neuron* 74:858–873.
- Welker WI (1964) Analysis of sniffing of the albino rat. *Behav* 22:223–244.
- Wesson DW, Wilson DA (2011) Sniffing out the contributions of the olfactory tubercle to the sense of smell: hedonics, sensory integration, and more?. *Neurosci Biobehav Rev* 35:655–668.
- Wesson DW, Donahou TN, Johnson MO, Wachowiak M (2008) Sniffing behavior of mice during performance in odor-guided tasks. *Chem Senses* 33:581–596.
- Wieland S, Schindler S, Huber C, Kohr G, Oswald MJ, Kelsch W (2015) Phasic dopamine modifies sensory-driven output of striatal neurons through synaptic plasticity. *J Neurosci* 35:9946–9956.
- Willmore B, Tolhurst DJ (2001) Characterizing the sparseness of neural codes. *Network* 12:255–270.
- Wilson DA, Sullivan RM (2011) Cortical processing of odor objects. *Neuron* 72:506–519.
- Xiong Q, Znamenskiy P, Zador AM (2015) Selective corticostriatal plasticity during acquisition of an auditory discrimination task. *Nature* 521:348–351.
- Yamaguchi M (2017) Functional sub-circuits of the olfactory system viewed from the olfactory bulb and the olfactory tubercle. *Front Neuroanat* 11:33.
- Zhang Z, Liu Q, Wen P, Zhang J, Rao X, Zhou Z, Zhang H, He X, Li J, Zhou Z, Xu X, Zhang X, Luo R, Lv G, Li H, Cao P, Wang L, Xu F (2017a) Activation of the dopaminergic pathway from VTA to the medial olfactory tubercle generates odor-preference and reward. *Elife* 6:e25423.
- Zhang Z, Zhang H, Wen P, Zhu X, Wang L, Liu Q, Wang J, He X, Wang H, Xu F (2017b) Whole-brain mapping of the inputs and outputs of the medial part of the olfactory tubercle. *Front Neural Circuits* 11:52.
- Znamenskiy P, Zador AM (2013) Corticostriatal neurons in auditory cortex drive decisions during auditory discrimination. *Nature* 497:482–485.