The Ingestive Response Reflects Neural Dynamics in Gustatory Cortex

Natasha Baas-Thomas[1*], Abuzar Mahmood[1*], Narendra Mukherjee[1], Kathleen C. Maigler[1], Yixi Wang[1], and Donald B. Katz[1,2]

Departments of Biology[1] and Psychology[2]

Brandeis University, Waltham, Massachusetts 02454, USA

[*] These authors contributed equally

Author Contributions:

NBT, AM, and DK conceived and designed the research. NBT, NM, KM, and YW performed the experiments. NBT, AM, and DK interpreted the results of the experiments. NBT and AM prepared figures. NBT, AM and DK drafted and revised the manuscript. All authors approved the final version of the manuscript.

Competing interest statement:

The authors declare no competing financial interests

Abstract

Upon delivery of a taste onto the tongue, the neural gustatory circuit must decide whether to ingest or reject the stimulus. While there has been much work in rodents investigating the neural activity leading to the rejection decision and its associated orofacial movements, characterization of behaviors related to ingestion have received much sparser attention. To address this, here we simultaneously measured electromyographic (EMG) activity of the jaw opener muscle and single-neuron activity from gustatory cortex (GC) so that we could characterize ingestion-related behaviors and their association with neural dynamics. We developed a machine-learning classifier to accurately identify distinct orofacial movements from EMG signals, thereby revealing multiple novel subtypes of ingestion-related behaviors. The frequency of occurrence of each shifts significantly around the time of the consumption decision, a change in behavior that is tightly coupled with the transition in GC's neural firing patterns into the state reflecting the tastant's palatability. These findings demonstrate a direct link between neural dynamics in GC and the orchestration of the physical movements that define ingestive behavior, highlighting GC's general role not just in taste perception and decision making, but also in the control of motor actions.

Introduction

When a taste stimulus reaches the tongue, the animal has one basic goal—that of determining whether that stimulus should be ingested or rejected. In rats, consumption judgments are actuated by a family of orofacial behaviors which are generated by overlapping sets of lingual and masticatory muscles. A multi-functional central pattern generator in the brainstem drives the full range of these behaviors (Z. Chen et al., 2001; DiNardo & Travers, 1997; Grill & Norgren, 1978b; Moriyama, 1987; J. Travers et al., 2000; Venugopal et al., 2007), but beyond this, much remains to be understood about the neural control of these consumption behaviors; in particular, little is known about how ingestion responses are modulated.

Tastant rejection has received more study, partly because the orofacial behavior that defines the decision to expel a tastant is a large and stereotyped (hence, easily-detected and classified) triangular-shaped opening of the mouth called a gape (Grill & Norgren, 1978a). The ease with which gapes are detected through electromyography (EMG) of the anterior digastric (AD) muscle (the jaw opener) has helped researchers relate rejection decisions to taste responses in the gustatory cortex (GC), which progress through three successive firing-rate "states", the last of which contains firing that is correlated with tastant palatability (Jones et al., 2007; Moran & Katz, 2014; Sadacca et al., 2012, 2016). The transition into this palatability state reliably precedes the onset of gapes to an aversive taste by ~0.3 s (Li et al., 2016; Sadacca et al., 2016), and brief (0.5s) optogenetic inhibition of GC delays gaping if (and only if) the onset of the silencing perturbation occurs before the transition into the palatability state (Mukherjee et al., 2019). In summary, the GC ensemble transition into palatability-related firing participates in driving of rejection behaviors.

Several lines of evidence support the suggestion that GC taste response dynamics also drive ingestive behaviors: firstly, GC palatability-related responses to palatable and unpalatable tastants are distinct but similarly timed (Katz et al., 2001; Lin et al., 2021; Sadacca et al., 2012); secondly, perturbation of GC affects discrimination of palatable tastes in both rodents (Bales et al., 2015) and humans (Pritchard et al., 1999); thirdly, GC projects to feeding centers (Saper,

1982) and parts of the brainstem (Shammah-Lagnado et al., 1992; C. Zhang et al., 2011; G. X. Zhang & Sasamoto, 1990) that control both ingestive and rejective orofacial movements (Berthoud & Münzberg, 2011; Z. Chen et al., 2001; Moriyama, 1987; J. Travers et al., 2000). This hypothesis has not been definitively tested, however, because the set of motor behaviors comprising ingestion is far more difficult to identify than gaping. In a seminal paper, Grill & Norgren (1978a) proposed that intraoral infusion of palatable tastants elicits a set of three distinct orofacial behaviors: lateral tongue movements (observed in response to only the most palatable tastants), mouth movements, and tongue protrusions. The latter two both involve rapid (~6-8 Hz), low amplitude jaw opening (Grill & Norgren, 1978a; J. B. Travers & Norgren, 1986) and, in the case of tongue protrusions, extension of the tongue beyond the incisor plane (Grill & Norgren, 1978a).

However, there is a lack of consensus with regard to the classification and function of these behaviors: some experimenters have suggested alternative definitions of "mouth movements" as movements of the lower mandible without mouth opening (Jarrett et al., 2005; López et al., 2023; Parker et al., 1992), while others (Kaplan et al., 1995) have suggested that mouth movements and tongue protrusions, rather than being categorically separate behaviors, instead exist on a single continuum of tongue extension; still others have suggested that mouth movements are hedonically neutral (Berridge, 1991; Brining et al., 1991; De Luca et al., 2012; Feurté et al., 2000), only weakly reflective of palatability at best (Berridge & Grill, 1983; Berridge & Treit, 1986; Kaplan et al., 1995), or not a part the ingestive profile at all (Fontanini & Katz, 2006; Pereira et al., 2021; Soares et al., 2007). To avoid this ambiguity, some experimenters eschew the intricacies of ingestive behaviors altogether, focusing on how neural circuits drive "licking" (broadly defined as any fast orofacial movement; Z. Chen et al., 2001; DiNardo & Travers, 1997; J. B. Travers & Jackson, 1992; J. B. Travers & Norgren, 1986).

Given this lack of clarity, testing the hypothesis that GC is involved in the driving of both ingestion and rejection will require (at minimum) a more sophisticated analysis of EMG responses. Here, we perform this analysis, using advanced data-mining approaches to systematically detect and categorize ingestive-related behaviors in EMG activity. Our use of information-rich EMG data mitigates the difficulty of distinguishing mouth movements from tongue protrusions in rats freely expressing consummatory behaviors without head restraint. The machine-learning classifier presented in this paper allows us to accurately discriminate individual orofacial movements using AD activity, and to thereby conduct a detailed analysis of ingestive-related behaviors, defined here as mouth or tongue movements (MTMs). This analysis revealed that MTMs are in fact made up of multiple distinct and novel behavioral subtypes—more subtypes than have hitherto been described. Further analysis demonstrated that the relative frequency of each subtype changes with the evolving GC response, and that transitions in behavior align with neural transitions of palatability-related firing.

These results demonstrate that consumption decisions are reflected in an ensemble of interacting behaviors, and provide further evidence that GC is involved not only in the generation of the rejection response, but also in behaviors related to ingestion. This further implicates GC in both processing gustatory stimuli and in effecting the subsequent motor response.

Methods

Experimental design and data collection

Subjects

Adult, female Long–Evans rats (*n* = 13, 250–400 g at time of surgical implantation, Charles River Laboratories) served as subjects in our study. A subset of these rats (n = 5) were taken from previously published data (Mukherjee et al., 2019; see **Table 1**). The rats were housed in individual cages in a temperature- and humidity-controlled environment under a 12 h light/dark cycle, given access to food and water *ad libitum* before the start of experimentation. All experimental methods were in compliance with National Institutes of Health guidelines and were approved in advance by the Brandeis University Institutional Animal Care and Use Committee.

Electrode, EMG, and intraoral cannula construction

Custom microwire bundle drives were constructed to include between 0 (for EMG-only implants) to 30 recording electrodes (0.0015-inch formvar-coated nichrome wire; AM Systems; see **Table 1**). The microwire bundle was affixed to a custom-made electrode interface board (San Francisco Circuits) and connected to a 32-channel Omnetics connector. The remaining two pins of the connector were wired to electromyography (EMG) electrodes (PFA-coated stainless steel wire; AM Systems) for muscle implantation. Finally, the microwires were connected to a custom-built 3D printed microdrive that allowed the entire assembly to be moved ventrally after implantation. For more information on the implanted apparatuses and associated electronics, see (Katz et al., 2001; Li et al., 2016; Sadacca et al., 2016), as well as the Katz Lab webpage (https://katzlab.squarespace.com/technology).

Intraoral cannulae, flexible polyethylene tubing (AM Systems) with a flanged tip and washer to ensure stability, connected to a plastic top complete with a locking mechanism, were built to allow the delivery of tastants directly onto the tongue as described previously (Fontanini & Katz, 2006).

Surgical implantation

All subjects were implanted with an intraoral cannula (IOC) and EMG in the anterior digastric; some additionally had an electrode implantation in GC. All rats were first anesthetized with an intraperitoneal injection of a ketamine/xylazine mixture (100 mg/kg and 5 mg/kg body weight, respectively), after which we shaved and cleaned the scalp and under the jaw, and situated the head in the stereotaxic frame. After excising the scalp and leveling the skull, we drilled five self-tapping screws into the skull for supporting and grounding the electrode bundles.

For subjects that had an electrode implant, we next made an additional, larger craniotomy (~2 mm dia.) above GC. An electrode microdrive (see *Electrode, EMG, and intraoral cannula construction*) was slowly lowered (over 5-10 min), and electrodes were positioned 0.5 mm above the GC (+1.3 mm A/P, +/- 5 mm M/L, 4.6 mm D/V). Afterwards, the craniotomy was sealed with Kwik-Sil (World Precision Instruments).

For all subjects, the ground wires were wound tightly around the skull screws and the microdrive was cemented in place with dental acrylic. The rat was then removed from the stereotaxic frame and implanted with a single (right-side) IOC in the space between the first maxillary molar and the lip, through the masseter muscle and inside the zygomatic arch, and out through the opening in the scalp (Katz et al., 2001; Phillips & Norgren, 1970) before being cemented in place.

The EMG electrodes were channeled down the left side of the face (opposite from the IOC); after the overlying skin had been teased away from the belly of the digastric muscle, one end of each EMG electrode was then inserted into the muscle using a suture needle (for more details, see Dinardo & Travers, 1994; Li et al., 2016; Loeb & Gans, 1986; Travers & Norgren, 1986). The electrode wires were trimmed and held in place with vetbond tissue adhesive (3M) and the skin covering the anterior digastric was sutured closed. Finally, bacitracin ointment was applied all around the base of the headcap and over the incision under the jaw.

Rats were postoperatively injected with analgesic (buprenorphine 0.05 mg/kg or meloxicam 0.04 mg/kg), saline, and antibiotic (Pro-Pen-G 150,000 U/kg). Similar antibiotic, saline and analgesic injections were delivered 24, 48 and 72 hr later, and bacitracin ointment was reapplied. The rats were allowed to recover for a minimum of 6 days, and daily weight records were kept to ensure that the rats did not fall below 82% of pre-surgery weight.

Experimental design

Habituation

Following recovery from the implantation surgery, we habituated rats to passive water deliveries in the experimental chamber for 2 to 3 days before beginning data collection. During these habituation sessions, 60 to 120 pulses of ~40 μL of distilled water were delivered into the animal's oral cavity through the IOC. We also placed rats on a mild water restriction schedule, 15-20 mL of water (not including the 4 mL delivered during habituation sessions themselves) per day to ensure engagement in the task, which was maintained for the duration of the study. At the end of the habituation period, electrode bundles were driven deeper such that the electrode tips lay within GC.

Passive taste administration

Experimental test sessions spanned 1 to 4 days after habituation (1 session per day). Freely moving rats received three to five tastants *via* IOC in a pseudo-random order. Each tastant battery (see **Table 1**), selected from the following set, included a range of distinct taste identities and palatabilities comparable to those in our (Grossman et al., 2008; Levitan et al., 2019; Moran & Katz, 2014; Sadacca et al., 2012) and others' (Samuelsen et al., 2012; Spector & Grill, 1988) studies: 0.1 M sodium chloride (NaCl), 0.3 M or 30 mM sucrose (Sucrose), 0.1 M citric acid (CA), 5 mM, 3 mM, 2 mM or 1 mM quinine-HCl (QHCl), 0.1 mM quinine-HCl (Dil-QHCl), or distilled water. Quinine-HCl concentrations (5-1 mM) and sucrose concentrations (0.3 M/30 mM) were each grouped, as these ranges elicit similar taste-responsive orofacial movements (Grill & Norgren, 1978a). During taste delivery, EMG and either video recordings or GC single unit activity were captured, as detailed below.

The taste delivery apparatus consisted of gently pressurized tubes containing taste solutions; the tubes converged upon a manifold of finer polyamide tubes that could be inserted

into (to 0.5 mm past the end of) the IOC, thus eliminating any chance of mixing. The manifold could be locked securely into the dental acrylic cap. The tastes were then delivered under slight nitrogen pressure—this taste delivery protocol has been consistently shown to ensure reliable tongue coverage at short latencies see (Katz et al., 2001; Li et al., 2016; Sadacca et al., 2016).

Acquisition of electrophysiological and EMG data

Electrophysiological signals from the microelectrodes were sampled at 30 kHz using 32-channel analog-to-digital converter chips (catalog #RHD2132) from Intan Technologies, digitized online at the head stage and sampled jointly, along with signals from actuators marking tastant delivery, using an Intan RHD USB interface board (catalog #C3100), which routed records to the hard drive of a PC for saving. The experimental chamber was ensconced in a Faraday cage that shielded recordings from external electromagnetic influences.

Video recording

A camera (Brio 4K, Logitech) was positioned below the plexiglass experimental chamber to best record the orofacial movements of the subjects. The videos were captured with OBS Studio to ensure that video settings were set to a resolution of at least 1920 x 1080 progressively displayed pixels at 60 frames per second. The delivery of a taste and a single red LED flash of 1 second were initiated simultaneously so that the light onset visually marked the start of the trial.

Histology

In preparation for histology to confirm GC electrode placement, rats were deeply anesthetized with an overdose of the ketamine/xylazine mixture. We perfused the rats through the heart with 0.9% saline followed by 10% formalin and harvested the brain. The brain tissue was incubated in a fixing mixture of 30% sucrose and 10% formalin for several days before being sectioned into 50 µm coronal slices on a sliding microtome (SM2010R, Leica Microsystems). Sections containing the electrode implant sites around GC were imaged with a fluorescence microscope (BX-Z710, Keyence). Histological verification for rats with only EMG implants was not necessary as targeting of EMG electrodes was directly observed during surgery.

Table 1. Summary of rat subjects and experimental conditions.

Rat ID	# of Sessions	Tastants	Trials / Tastant	Data Types			Published
				Anterior Digastric EMG	Video	# of GC Electrodes	
NB27	3	0.3M Suc, 0.1M NaCl, 0.1M CA, 5mM QHCl	30	✓	✓	✗	✗
NB32	3	Water, 0.3M Suc, 0.1M NaCl, 2mM QHCl	30	✓	✓	✗	✗
NB34	3	Water, 0.3M Suc, 0.1M NaCl, 3mM QHCl	30	✓	✓	✗	✗
NB35	2	Water, 0.3M Suc, 0.1M NaCl, 2mM QHCl	30	✓	✓	✗	✗
KM29	2	0.3 Suc, 0.1M NaCl, 0.1M CA, 1mM QHCl	30	✓	✗	14	✗

KM45	1	Water, 0.3M Suc, 0.1M NaCl, 0.1M CA, 5mM QHCl	25	✓	✗	16	✗
KM50	2	Water, 0.3M Suc, 0.1M NaCl, 0.1M CA, 5mM QHCl,	25	✓	✗	16	✗
NB33	1	Water, 0.3M Suc, 0.1M NaCl	40	✓	✗	30	✗
NB33	1	Water, 0.3M Suc, 0.1M NaCl, 2mM QHCl	30	✓	✗	30	✗
NM51	1	0.3M Suc, 30mM Suc, 0.1mM QHCl, 1mM QHCl	8	✓	✗	30	✓
NM51	1	0.3M Suc, 30mM Suc, 0.1mM QHCl, 1mM QHCl	15	✓	✗	30	✓
NM50	2	0.3M Suc, 30mM Suc, 0.1mM QHCl, 1mM QHCl	8	✓	✗	30	✓
NM50	2	0.3M Suc, 30mM Suc, 0.1mM QHCl, 1mM QHCl	15	✓	✗	30	✓
NM47	1	0.3M Suc, 30mM Suc, 0.1mM QHCl, 1mM QHCl	15	✓	✗	30	✓
NM47	2	0.3M Suc, 30mM Suc, 0.1mM QHCl, 1mM QHCl	8	✓	✗	30	✓
NM44	1	0.3M Suc, 30mM Suc, 0.1mM QHCl, 1mM QHCl	8	✓	✗	30	✓
NM44	1	0.3M Suc, 30mM Suc, 0.1mM QHCl, 1mM QHCl	10	✓	✗	30	✓
NM43	2	0.3M Suc, 30mM Suc, 0.1mM QHCl, 1mM QHCl	8	✓	✗	30	✓
NM43	1	0.3M Suc, 30mM Suc, 0.1mM QHCl, 1mM QHCl	10	✓	✗	30	✓
NM43	1	0.3M Suc, 30mM Suc, 0.1mM QHCl, 1mM QHCl	15	✓	✗	30	✓

Data processing and statistical analysis

The analysis of data and statistical tests were performed using custom-written software in Python as described below. Some of the more prominently used packages are mentioned here:

- Scipy (statistical analysis; Virtanen et al., 2020)
- Pandas (data handling and manipulation; The pandas development team, 2025)
- Scikit-Learn (machine learning; Pedregosa et al., 2011)
- Numpy (numerical computing; Harris et al., 2020)
- Pingouin (hypothesis testing; Vallat, 2018)
- Blech_clust (electrophysiological data processing; Mahmood et al., 2025)
- Scikit-Optimize (hyperparameter optimization; Head et al., 2024)
- UMAP (non-linear dimensionality reduction; McInnes et al., 2018)

Video scoring

Scoring methodology and ethogram

Videos of each session were first reviewed to identify instances in which the rat's mouth was clearly visible for a few (5.1±3.7 s) seconds following taste delivery. Given that the rats are freely moving, only a fraction of the total trials was usable for labeling (13% of all trials) due to obstructed views from odd head angles or forepaws resting in front of the mouth in most trials. The selected trials were then analyzed frame-by-frame to classify and record the subject's

orofacial movements using BORIS (Friard & Gamba, 2016). All scoring was performed blind to the delivered taste to minimize potential bias. Two scorers independently evaluated the presence of three taste-responsive orofacial movements: gape (retraction of the corners of the mouth, resulting in a wide triangle-shaped opening of the mouth), mouth and togue movements (MTMs, rhythmic, low-amplitude openings of the mandible, with or without tongue emergence), and lateral tongue protrusion (lateral extension of the tongue). Note that mouth movements (rhythmic, low-amplitude openings of the mandible) and tongue protrusion (emergence of the tip of the tongue directly on the midline during mouth opening) are combined into one behavior: MTM. This decision was made because, in our recordings, it was not possible to reliably distinguish whether or to what extent the tongue emerged, making it difficult to accurately separate these behaviors. Furthermore, given the lack of consensus in the literature regarding the distinction between these two behaviors (see Introduction), we felt this decision was well justified. Periods of no movement (absence of any perceptible orofacial muscle activity) were also marked by the scorers, though this category was not assessed for inter-rater accuracy due to its simplicity.

Since video sections were focused on taste responsive sections, we supplemented the number of "no movement" samples by using the EMG signal to select pre-stimulus (1000-2000 ms pre-stimulus) activity trials with no discernible orofacial movements (based on the amplitude of the pre-stimulus signal relative to the average post-stimulus signal amplitude) and incorporating all segmented behaviors from that time period as "no movement" samples.

Interrater scoring accuracy

To assess inter-rater accuracy, we analyzed the first 5 seconds of trials that were independently scored by both observers. In each comparison, one scorer's behavior annotations were treated as the "ground-truth," and we evaluated whether the second scorer had a temporally overlapping annotation of the same behavior type. Any overlap in annotation was considered an agreement; annotations without any overlap were considered disagreements. Accuracy was computed for each behavior category (MTMs, gapes, LTMs). This analysis was then repeated with the second scorer treated as the ground truth. Agreement accuracy was then calculated as the proportion of matched bouts out of the total number of bouts for each scorer. This can be formalized as:

$$\text{Accuracy}_i = \frac{1}{j}\sum^{j} \psi_k\left(x_j\left(t_j, \Delta t_j\right)\right)$$

$$\psi_k = \begin{cases} 1 \text{ if } \max(x_k \odot x_{k`}) = 1 \\ \text{else } 0 \end{cases}$$

$i =$ Counter over label categories

$j =$ Counter over label instances for a given labeler

$k =$ Counter over labelers

$x\left(t_j, \Delta t_j\right)$ is a binary vector representing the duration of a specific labelled event

To estimate a chance level of agreement, we repeated the analysis 100 times with randomized annotations. Specifically, in each shuffle iteration, the behavior annotations of a scorer were randomly reassigned new labels (uniformly sampled across MTM, gape, and LTM), after which, agreement accuracy was recalculated.

Electrophysiology

Single unit isolation

Spikes from electrophysiological recordings were sorted and analyzed off-line using an in-house Python-based pipeline (Mahmood et al., 2025). Putative single-neuron waveforms with >5:1 signal-to-noise ratio (median absolute deviation) were sorted using a semi-supervised algorithm: recorded voltage data were filtered between 300-3000 Hz, grouped into potential clusters by a Gaussian Mixture Models (GMM) fit to multiple waveform features; clusters were then labeled and/or refined manually (to increase conservatism) by the experimenters.

Electromyography

EMG processing

As detailed previously, we recorded voltage signals from two unipolar EMG electrodes implanted in the anterior digastric muscle at 30 kHz. We used the difference in the voltage recorded by the two electrodes as the EMG signal—this procedure helps to cancel out any large artifacts produced by the animal's movements and is equivalent to using a differential amplifier (as done in Li et al., 2016). We down-sampled the EMG signal to 1000 Hz by averaging the voltage values in sets of 30 samples, and high pass filtered the down-sampled signal above 300 Hz (Li et al., 2016; J. B. Travers & Norgren, 1986) using a 2nd order Butterworth filter. The absolute value/magnitude of the filtered EMG signal was then lowpass filtered (again using a Butterworth filter of order 2) below 15 Hz, effectively capturing the envelope of variation of the EMG signal (plotted as the black curve in **Figure 1A**). This cutoff of 15 Hz is sufficient for identifying orofacial behaviors, all of which occur at frequencies smaller than 10 Hz (Li et al., 2016; J. B. Travers & Norgren, 1986).

XGB Classifier

Code for generating and running the classifier, as well as the final training dataset can be found at https://katzlabbrandeis/blech_emg_classifier

Movement extraction and refinement

We used low-pass filtered EMG signals (as described above) to extract individual orofacial movements by first applying a top-hat filter (scipy, length = 200 ms) to the signal and segmenting contiguous non-zero regions of the filtered signal. In some cases, this resulted in waveforms which had biologically implausible durations (too long or short). We removed any extracted waveforms shorter than 50 ms and longer than 500 ms.

Removal of multimodal MTM waveforms

We found that our process extracted some MTMs waveforms that were multimodal (3.9% of all MTM waveforms), likely due to two muscle contractions that were improperly segmented. Despite their small fraction, these waveforms added substantial variability to the data and therefore needed to be removed. To assess whether individual MTM waveforms were unimodal or multimodal, we correlated the non-normalized waveform with itself (scipy.signal.correlate), retaining only the positive lags (positive lagged autocorrelation). The presence of one or more peaks (identified using scipy.signal.find_peaks) indicated that the

waveform contained repeating structure consistent with multimodality. Waveforms identified as multimodal were removed from the dataset and from all subsequent analyses.

Expansion of no-movement category using NCA and GMM

Due to the noisy nature of EMG activity, as well as the imperfect labelling and extraction of individual orofacial movements, we observed taste-related orofacial movements (mostly MTMs but also some gapes) with very small amplitudes, likely due to the unavoidable labelling of all orofacial movements within a range of time from video scoring. To address this issue, we absorbed / relabeled mouth movements most similar to the "no movement" category in feature space. To do this, we performed Neighborhood Component Analysis (NCA) to extract the 2 dimensions which following a linear transformation best allow classification of all 3 mouth movement categories using a Nearest Neighbor Classifier. In the 2D-NCA space, we fit Gaussian distribution to the "no movement" category and reclassified all mouth movements which were within 2 Mahalanobis distances of this Gaussian.

Feature engineering

We extracted a total of 8 features for each individual mouth movement. Two of these—duration and time to next peak—were based on the previously published QDA classifier (Li et al., 2016), and another—maximum frequency—was based on the previously published BSA classifier (Mukherjee et al., 2019). We included 5 additional features—amplitude, time to previous peak, and the first 3 Principal Components of the amplitude and duration normalized waveforms—to assess whether these improved classification performance. PCA was trained using waveforms from the entire training dataset. All features were then normalized using the StandardScaler implemented in Scipy. Such extraction of features is standard practice to improve computational efficiency and model fitting, especially for clustering algorithms by reducing the dimensions representing each datapoint (Khalid et al., 2014). Reducing the dimensionality of the data combinatorially reduced computational complexity and decreases the sparsity of representational space of the data, allowing clustering algorithms to more effectively discover spatial patterns (Radovanović et al., 2010). Both the PCA and StandardScaler objects were then frozen (dumped) for future processing and available for reuse from the GitHub repository.

Hyperparameter optimization

To ensure robustness of our classifier, we tested that the training of our classifier was agnostic to a specific set of hyperparameters used. We used the scikit-optimize library (Head et al., 2024) to perform 20 optimization runs with 400 iterations/run using leave-one-animal-out cross-validation accuracy as the optimization target, and the "GBRT" model as a surrogate function to sequentially model the fitness landscape (**Table 2**). The chosen set of hyperparameters is available in the GitHub repository.

Table 2. Optimization space of classifier hyperparameters.

Name	Domain	Distribution	Min	Max
learning_rate	Real	log-uniform	0.001	1.0
min_child_weight	Integer	Discrete uniform	1	10

max_depth	Integer	Discrete uniform	1	100
max_delta_step	Integer	Discrete uniform	0	20
subsample	Real	Uniform	0.01	1.0
colsample_bytree	Real	Uniform	0.01	1.0
colsample_bylevel	Real	Uniform	0.01	1.0
reg_lambda	Real	log-uniform	1e-9	1000
reg_alpha	Real	log-uniform	1e-9	1.0
gamma	Real	log-uniform	1e-9	0.5

Determining important features

We assessed the importance of features using SHAP analysis (Lundberg et al., 2020). Briefly, SHAP (SHapley Additive exPlanations) is a game theoretic approach to explain the output of any machine learning model. It connects optimal credit allocation with local explanations using the classic Shapley values from game theory and their related extensions. We used the Mean Absolute SHAP values implemented in the SHAP package (https://github.com/shap/shap) as a metric for feature importance.

Evaluation Classifier Performance

In order to quantitatively evaluate the performance of our classification model, we employed the standard metric of Accuracy, along with three additional metrics: precision, recall, and F1-score to obtain a more detailed comparison of the models.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Where TP=True Positive, FP = False Positive, FN = False Negative, FP = False Positive

Precision (also known as Positive Predictive Value) assesses the model's accuracy among its positive predictions, effectively answering the question "Of all the events the model identified as positive, how many were actually positive?". A high precision value indicates a low rate of false positives, which is critical in applications where such errors are costly.

Recall (also known as Sensitivity) measures the model's ability to find all the positive events within the data, addressing the question "Of all the actual positive events in the data, how many did the model correctly identify?". A high recall value indicates a low rate of false negatives, which is particularly important in screening applications where missing a positive case is undesirable. Given the inherent trade-off between precision and recall, a model optimized for

high precision may miss many true positives, while a model optimized for high recall may generate many false positives.

To provide a single, balanced metric that accounts for this trade-off, we calculated the F1-score. The F1-score is the harmonic mean of precision and recall, providing a more robust measure of overall model performance, especially in datasets with an imbalanced class distribution. This score ranges from 0 to 1, with a value of 1 representing perfect precision and recall. The F1-score is therefore a suitable metric for evaluating a model where both false positives and false negatives are of significant concern.

We note precision, recall, and F1 score overlap with the commonly used d' (Sensitivity Index or Fisher Discriminant Index) which is an unbiased estimator of the capacity of a classifier for binary classification.

Specificity to post-stimulus period

To assess consistency of model predictions to the taste-responsive period (previously shown to be primarily between 0-2000 ms post-stimulus delivery (Katz et al., 2001; Sadacca et al., 2012), we calculate for gapes and MTMs the fraction of events observed between 0-2000 ms post-stimulus delivery out of the total prediction window of -1000-3000 ms post-stimulus delivery.

Analysis of mouth or tongue movements (MTMs)

Divergence analysis of MTM feature distributions

For each session, MTM feature vectors were z-score normalized prior to dimensionality reduction using Uniform Manifold Approximation and Projection (UMAP). Each MTM was labeled as occurring either "before" or "after" 800 ms post-stimulus delivery, the average onset of GC palatability-related firing (Katz et al., 2001; Mukherjee et al., 2019; Sadacca et al., 2012, 2016). To quantify differences in the spatial distribution of MTMs between the two event positions, a histogram difference matrix was obtained by subtracting the binned density values of the "after" group from those of the "before" group. A divergence statistic was defined as the sum of the absolute values of all non-zero overlapping bins in this difference matrix.

Given the stochasticity of the embedding generated by UMAP, to assess the statistical significance of the observed divergence, we conducted a permutation-based test, repeating the test 10 times to ensure the robustness of our results. For each repeat, 1,000 bootstrap samples were generated by randomly assigning MTM events to "before" and "after" groups while preserving group sizes. For each shuffled comparison, a divergence statistic was computed as described above. P-values were computed as the proportion of bootstrap statistics exceeding the observed divergence.

Clustering MTMs into behavioral subtypes

MTMs that were classified by the classifier were analyzed on a session-by-session basis. The scaled features of the MTMs were first reduced to 90% using principal component analysis. These were then clustered using Gaussian Mixture Model (GMM) and the Bayesian Information Criterion (BIC) scores of between 1 to 14 clusters were calculated. The optimal BIC score was

determined by finding the "elbow", i.e. the single breakpoint of a piecewise linear regression. This process was repeated for 50 iterations for each session.

As the mean optimal cluster number was 3, we subsequently repeated the process above, but set the GMM to a fixed 3 clusters for all sessions, such that every MTM had an associated cluster label of either 1, 2, or 3. We verified that the clusters were separable by calculating the pairwise Mahalanobis distances between all clusters, and confirmed that inter-cluster distances were significantly larger than intra-cluster distances using a Kolmogorov-Smirnov test.

Alignment of MTMs cluster labels across sessions

We aligned the labels across test sessions because cluster index labels are randomly assigned during the clustering process above. We performed this alignment using a "greedy" (sequential) approach. We first calculated the average feature vector for each cluster from two sessions and computed the cosine similarity between each pair. Cluster labels were then reordered to maximize the summed cosine similarity. The three clusters were then merged across sessions, and the resulting clusters were compared to the next session. This process was repeated, comparing the merged set of clusters to a new session, until all sessions were aligned.

$$Cos(\theta_{A,B}) = \frac{\vec{A} \cdot \vec{B}}{||\vec{A}|| \; ||\vec{B}||}$$

$$\text{Summed Cosine Similarity} = \sum_{i,j} Cos(\theta_{i,j}) \;\; \forall \, i \neq j$$

Support vector machine verification of MTM subtype labels

As a validation that the MTM clusters were separable, we trained a support vector machine with a radial basis function kernel (scikit-learn.svm) on the 8 waveform features (see "Feature engineering", Methods) using a leave-one-session-out cross-validation scheme. For each iteration, one session was designated as the test set, and all data from the corresponding animal were excluded from the training set. Predictions were generated on the held-out session, and classification accuracy was computed. To assess whether accuracy exceeded chance (33%), a one-tailed t-test was performed.

Analysis of MTM waveform feature differences across clusters

We extracted seven additional features from each non-normalized waveform segment to further characterize how waveform shapes differed between clusters: amplitude, area under the curve, full width half maximum, symmetry, skew, rising slope, and falling slope. Of note, symmetry was quantified as the Pearson correlation between the waveform and its mirrored version, and skewness was calculated using the sample skew statistic (scipy.stats.skew). Feature differences across clusters were evaluated with the Kruskal-Wallis test, and significant features with substantial effect sizes were followed up with a Mann-Whitney U-Test.

Temporal analysis of behavioral event counts

Gapes and the three MTM subtypes between 0.3 s to 1.3 s post-stimulus delivery (a 1 s window around the average time to palatability-related firing in GC) were isolated for every trial. Behaviors were counted as either "before" or "after" the 800 ms mark. For each test session, a Poisson means test was used to determine whether movement counts differed significantly between the "before" and "after" groups for every taste. To assess whether the number of significant changes collectively exceeded the proportion expected by chance with an alpha of 0.05 (5%), we generated a null distribution of p-values using a shuffling procedure. A binomial test confirmed that the number of significant p-values observed was significantly greater than expected by chance.

Changepoint estimation

We used a Categorical changepoint model implemented in the pytau package (https://github.com/abuzarmahmood/pytau) to infer when changes in the probability of occurrence happened in the timeseries of clustered oral behaviors. Models were fit using Automatic Differentiation Variational Inference (ADVI) functionality in the pymc library for 10e5 iterations (Salvatier et al., 2016) . Briefly, the structure of the model is given below. The subscripts following the distribution denote the size of the variable, where each individual element is an independent variable.

$$\alpha_{tau} \sim \text{HalfCauchy}(3)_{n-changes}$$

$$\beta_{tau} \sim \text{HalfCauchy}(3)_{n-changes}$$

$$\tau_{latent} \sim \text{Beta}(\alpha_{tau}, \beta_{tau})_{(trials,n-changes)}$$

$$\lambda \sim f(\tau_{latent}, p)_{(n-trials,time,n-behaviors)}$$

$$Obs \sim \text{Categorical}(\lambda)_{(n-trials,time,n-behaviors)}$$

For p, emissions for each state are modelled by an independent Dirichlet distribution, where each distribution is initialized with an even probability of 1 across all behaviors (1n-behaviors)
n-changes = n-states - 1 (both are used for clarity, above)
f = A function which combines latent and p to generate the inferred probability for each behavior for each trial and timestep.

Determining best number of states for MTM data

We fit changepoint models with 2-8 states to all datasets using ADVI as described above. ADVI optimizes the Evidence Lower Bound (ELBO), which is an estimate of the marginal likelihood of the model, providing an estimate of goodness of fit while accounting for model complexity (number of parameters), in a similar way to frequently used information criteria as AIC and BIC (Akaike, 1998; Dziak et al., 2019).

Inferring change in neural palatability correlation

We calculated the timeseries of palatability correlation for neurons recorded in GC by performing a Spearman correlation of the moving-window firing rate of each unit in response to every tastant with the canonical palatability rank of that tastant. The canonical palatability rank for each tastant was determined using the volume of the consumed in a Brief Access Task. We performed Principal Component Analysis on the set of palatability correlations across all neurons to extract the timeseries best representing the dynamics for the dataset (PC 1). To infer the largest change in this PC 1, and also obtain an estimate of the uncertainty of our inference, we fit the above timeseries with a Bayesian sigmoid model using ADVI for 10e5 iterations. We used broad hyperpriors, as is standard practice, to avoid over-constraining the model. The model structure is as follows:

$f(a,b,c,d,t) = \frac{a}{1+\exp(-b*(t-c))} - d$, where t denotes time, and a,b,c,d are parameters.

$a,b,d \sim \text{HalfNormal}(\sigma=5)$

$c \sim \text{Uniform}(0, t_{max})$

$\nu \sim \text{Exp}(\lambda = 1)$

$\sigma \sim \text{HalfNormal}(\sigma=1)$

$Obs \sim \text{StudentT}(\mu = f(a,b,c,d,t), \nu, \sigma)$

Monte-Carlo simulation to test for random alignment

We calculated the probability of obtaining an absolute distance between the first MTM changepoint and the palatability inflection point by sampling uniformly from the space $(t_1, t_2) \in t_1, t_2 < t_{max}; t_1 < t_2$, and calculating the minimum distance between the palatability inflection point and t_1, t_2 to provide a null-distribution for distances that can be obtained by chance. We calculated the percentile of the actual distance to this null-distribution to obtain a p-value.

Estimating effect of removing gapes and nothing movements on state similarities

We reasoned that if gapes and nothing movements were driving state changes, then removing them would significantly decrease the dissimilarity of subsequent states. To perform this calculation, we used the mean occurrence probabilities of behaviors in each state inferred by the changepoint model, defined as follows:

$\vec{p} \sim \text{Dirichlet}(\vec{\alpha}), \alpha \in [0, \text{inf})$

As a measure of dissimilarity, we calculated the norm of the difference in probability vectors between subsequent states.

$\left|\left|\overrightarrow{\Delta p}\right|\right| = \left|\left|\vec{p}_s - \vec{p}_{s+1}\right|\right|_2 : s \in \{0,1\}$ as there is a total of 3 states.

We standardized the distribution of these $\left|\left|\overrightarrow{\Delta p}\right|\right|$ for all behaviors and for only MTMs using the mean norm difference expected by random vectors of corresponding sizes to account for the difference in the dimensionality of the vectors.

Data availability

We have structured our electrophysiology datasets in a hierarchical data format (HDF5) and are hosting the files on a university-wide network share managed by Library and Technology Services (LTS) at Brandeis University. These HDF5 files contain our electrophysiology recordings, sorted spikes, and single-neuron and population-level analyses (and associated plots and results). These files are prohibitively large to be hosted on a general-purpose file share platform; we request people who are interested in our datasets to contact the corresponding author, Donald Katz (dbkatz@brandeis.edu), who can put them in touch with LTS to create a guest account at Brandeis University to securely access our datasets (hosted on the katz-lab share at files.brandeis.edu). As noted above, training data for the classifier is available as part of the GitHub repository.

Results

Overview

Our investigation into whether ingestive-related orofacial movements change in relation to GC dynamics occurs in a sequence of 3 main parts.

First, we developed a machine-learning classifier that enabled us to identify individual orofacial movements from AD EMG activity. We began by labeling EMG waveforms as either gapes, MTMs, or no movement, using scored videos. A gradient boosted tree-based classifier was then trained on features extracted from these labeled waveforms. We validated the classifier's accuracy, ensuring it can perform well on out-of-sample (novel) data, and confirmed that classifier performance matched or exceeded previously published gape detection algorithms (Li et al., 2016; Mukherjee et al., 2019).

Second, having verified that we could reliably classify both ingestive and aversive orofacial movements, we conducted a detailed analysis of MTMs using labels inferred by the classifier. This analysis revealed MTMs to be comprised of three distinct behavioral subtypes and showed that the overall structure of MTMs shifts over the course of a single trial, driven by frequency changes across all three MTM subtypes. The use of EMG further enabled us to resolve subtle differences in the properties of these behavioral subtypes that would be difficult to distinguish from video.

Lastly, we sought to determine whether changes in occurrence of MTM subtypes align with decision-related GC dynamics, as has been shown for rejective behaviors. Using data from test sessions in which ensembles of GC neurons were recorded simultaneously with AD EMG activity, we showed that changes in MTM subtype frequencies are aligned to the onset of palatability-related firing in GC. This result provides evidence that GC is involved in initiation of both ingestive and rejective behaviors.

Video scoring provides a training dataset of individual AD EMG waveforms

In order to reliably assess ingestion-related mouth movements, we first needed to develop a machine learning classifier capable of discriminating ingestive and rejective behaviors using AD (jaw opener) EMG activity. Toward this end, we generated a training dataset in which taste-responsive behaviors identified from videos by expert scorers were applied to AD EMG waveforms (see **Fig. 1A** for experimental setup).

This training dataset was comprised of nine sessions in which freely moving rats (n = 4) were presented with a battery of basic tastants (0.3 M Sucrose, 5 mM Quinine, 0.1 M Citric Acid, and 0.1 M Sodium Chloride) *via* an implanted intraoral cannula. Video recordings (sampling rate = 60 Hz) and AD EMG activity simultaneously captured (**Fig. 1A**) orofacial responses to the tastes. Two independent video scorers, both blind to tastant identity, evaluated the subset of trials in which the rat's mouth was clearly visible (13% of the total number of trials). The scorers labeled time points (5.1±3.7 s of post-stimulus activity, mean±SD) at which consummatory-related behaviors could be identified. The ethogram used for scoring followed mimetic descriptions established by Grill and Norgren (1978a), with one modification: mouth movements and tongue protrusions were combined into a single behavioral category, "mouth or tongue movements (MTMs)", which were defined as rhythmic, low-amplitude openings of the mandible, with or without tongue emergence (see Methods for explanation of this decision).

Consistency of video labels from the two scorers was assessed using an accuracy classification score which accounts for the temporal component of video labels (a component that other standard inter-rater reliability metrics typically neglect). One scorer was designated as the reference ("ground-truth") and the other's scores were compared against it; the designations were then reversed and the resulting mean accuracy value per behavior was compared to a null distribution generating by shuffling labels. Similarity of labels from both scorers was significantly higher than expected by chance (all p-values < 0.001, one-tailed t-test) for all three behaviors (MTMs: accuracy = 0.712, shuffle = 0.50±0.012; gapes: accuracy = 0.772, shuffled = 0.42±0.016; LTMs: accuracy = 0.735, shuffle = 0.09±0.005; mean±SD, **Fig. 1B**), demonstrating that the scorers were correctly classifying the behaviors.

Time-synced EMG records of AD activity were then aligned to the video record, such that time segments corresponding to the user-scored behavior labels could be matched to waveforms extracted from the EMG signal. **Figure 1C** presents examples of this analysis—time plots of behavioral events that occurred within 2 seconds of taste deliveries, along with the corresponding EMG activity. After applying a filter to subtract the rolling baseline EMG activity (see Methods), individual waveforms were extracted as all continuous non-zero regions and subsequently grouped according to their associated behavior label (**Figure 1D**). Each peak is identified as a single orofacial movement as done previously (Li et al., 2016; Mukherjee et al., 2019). MTMs comprised a large majority of these behaviors (66.1% of a total n = 2,176) while gapes constituted 14.3%. We also isolated an equivalent number of "no movement" waveforms (14.8%; **Figure 1E**). As there were, as expected, very few identified LTM waveforms (4.7%)—too few for us to train as a distinct category for classifier training—these were excluded from further analysis.

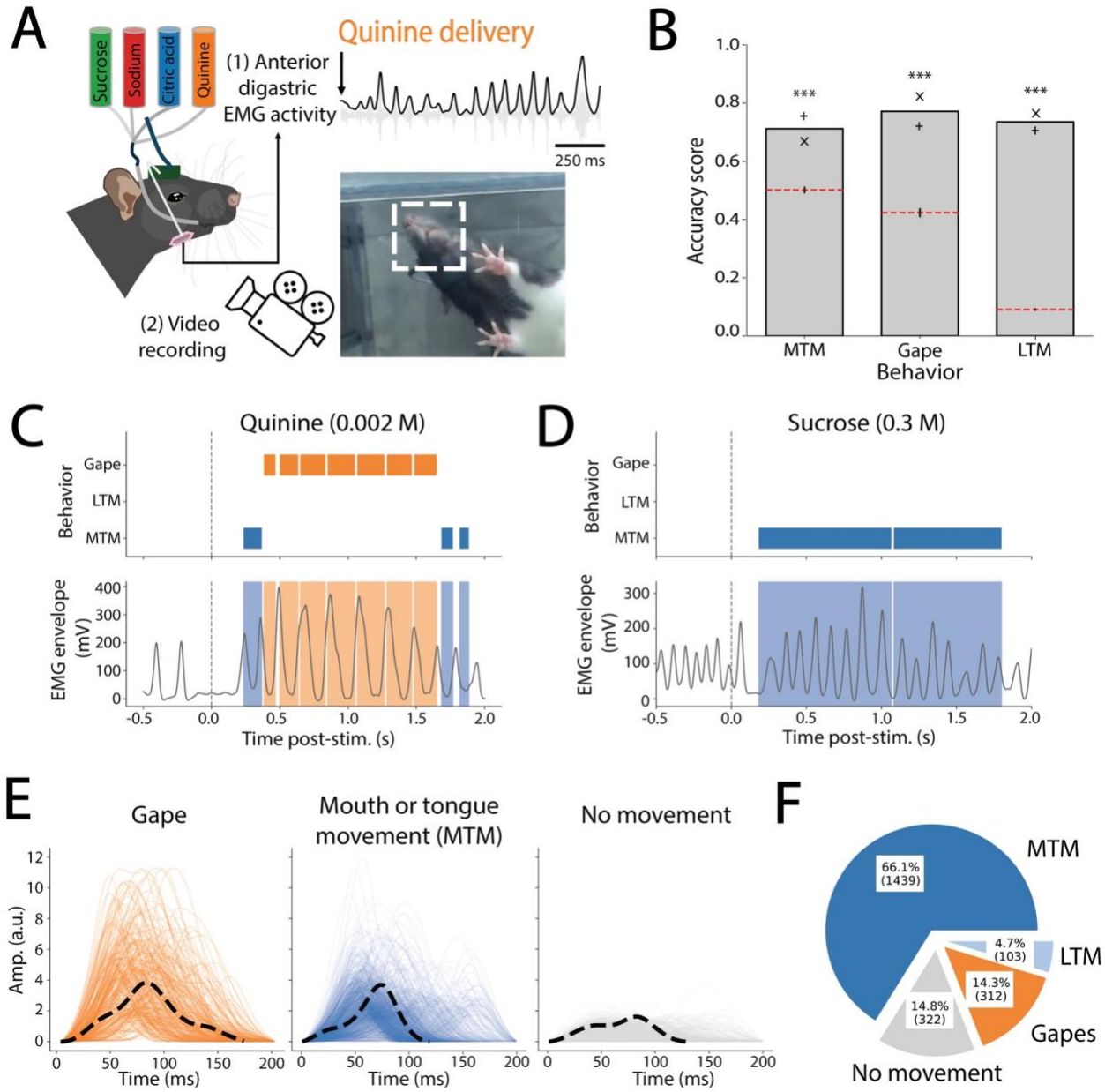

Figure 1. Annotation of EMG waveforms using scored video labels. A) A schematic of the data recording and annotation setup. Rats implanted with electromyography (EMG) electrodes and an intraoral cannula were given a battery of tastants while EMG and video recording were captured. **B)** Video scoring similarity between the two scorers was significantly higher than chance for all labels (*** $p < 0.001$, one-tailed t-test). x = accuracy score with scorer 1 as ground-truth, + = scorer 2 as ground-truth. **C-D**) Examples of video-scored behaviors (top row), and corresponding EMG signals (bottom row) for trials of quinine (C) and sucrose (D). **E)** Overlaid waveforms for each group of labelled movement (Note: Lateral Tongue Movements [LTMs] are not shown due to having too few waveforms). **F)** Breakdown of the labelled dataset by behavior

A machine-learning classifier accurately captures multiple classes of orofacial movements in EMG signals

To infer behavioral responses from EMG traces, previous studies have either performed visual analysis (Dinardo & Travers, 1994; Sasamoto et al., 2001; J. B. Travers & Norgren, 1986), used classifiers to assign labels to individual orofacial movements (Quadratic Discriminant Analysis [QDA], Li et al. 2016) or identified the probability of a rhythmic train of events (Bayesian Spectrum Analysis [BSA], Mukherjee et al. 2019) using solely the frequency of the signal. Each approach has allowed for successful investigations into the relationship between neural activity and rejective behavioral outputs, with classification algorithms further facilitating the automation of this process. However, both classifiers were trained and validated purely for the prediction of gapes (Li et al., 2016; Mukherjee et al., 2019), and are therefore unsuitable for classification of ingestion-related movements (i.e., MTMs). Put simply, these previous algorithms performed well because gaping is easily isolated and already well-understood behavior that an algorithm can be trained to specifically detect, and are unlikely to perform well for other, more subtle behaviors (as we demonstrate below).

To address these limitations, we developed a more complex classification model based on a gradient-boosted trees algorithm (XGB, based on the XGBoost algorithm, Chen & Guestrin, 2016) capable of performing classification of the extended set of the labelled EMG signals of orofacial movements. Rather than directly training the classifier on the EMG waveforms, we first extracted a suite of characteristics (features) from each waveform, a decision that improved the computational efficiency of the classifier and allowed us to interpret the differences in features between differently classified orofacial movements (in line with previous work, see Li et al., 2016; Mukherjee et al., 2019; see also Methods for a discussion on how such "feature engineering" is a standard way to address dimensionality issues). By applying this method we make use, in a unified and principled framework, of the range of information sources that have been used in the various classic forms of EMG analysis.

The use of a high-capacity tree-based model provides excellent interpretability (**Fig. 2A** shows a schematized version of the algorithm as a mixture of trees) and the opportunity to explore whether incorporating features beyond those previously used in classification algorithms (see below) improves classification accuracy. We extracted the following features: amplitude, multiple frequency-based properties including maximum frequency, time from previous peak, time to following peak, and duration, and shape-based properties extracted using the first three principal components of standardized waveforms from the training set (see Methods for more details). We then applied SHAP-based model interrogation (**SH**apley **A**dditive ex**P**lanations, Lundberg et al. 2020) to test the contribution of each feature to the predictions of the model. This analysis confirmed that each of the eight features contributed significantly compared to a control (Null SHAP values) generated by models trained on shuffled labels ($p < 0.001$ for all comparisons of actual vs null SHAP values, Wilcoxon Signed-Rank Tests). Since SHAP values are additive, we obtained a better understanding of the importance of each feature group by visualizing the summed SHAP-values for each group of features. We found that amplitude is the most important feature, followed by frequency- and shape-related features. (**Fig. 2B**). We, however, found no significant differences between the importance of features within each group (i.e. all frequency features had equal importance to each other, as did all shape features, $p > 0.1$ for features in frequency group, and in shape group, ANOVA).

Given that the model was initialized with specific hyperparameters (parameters determining model structure and training that do not get optimized as part of training), we tested whether the model's performance was dependent on those hyperparameters by performing 20 repeats of sequential optimization, randomly selecting starting hyperparameters anew with each repeat and optimizing for 400 iterations (in **Fig. 2C**, each repeat is represented by a different color). We found that the leave-one-animal-out cross-validation accuracy of the model, which is high with the initial random set of hyperparameters (accuracy = 0.74±0.02, mean±sd, compare with chance accuracy = 0.33), increases only slightly following hyperparameter optimization (~4% increase, 0.78±0.002, $p < 0.001$, paired t-test versus initial values). Following this optimization, the distribution of hyperparameters from the 20 runs covered almost the entire search range for each hyperparameter; nonetheless, the model converged to similar values of cross-validation accuracy (see Methods for search space specifications, fraction of search space covered by final distributions = 0.93±0.05, mean±sd across hyperparameters)—that is, hyperparameter sets with widely varying values gave similarly high accuracies (although some hyperparameters did show tighter clustering, **Fig. 2C Inset**). Overall, this result demonstrates that the predictive capacity of the model reflects robust interpretability of the EMG data, rather than some idiosyncrasy of model definition and training.

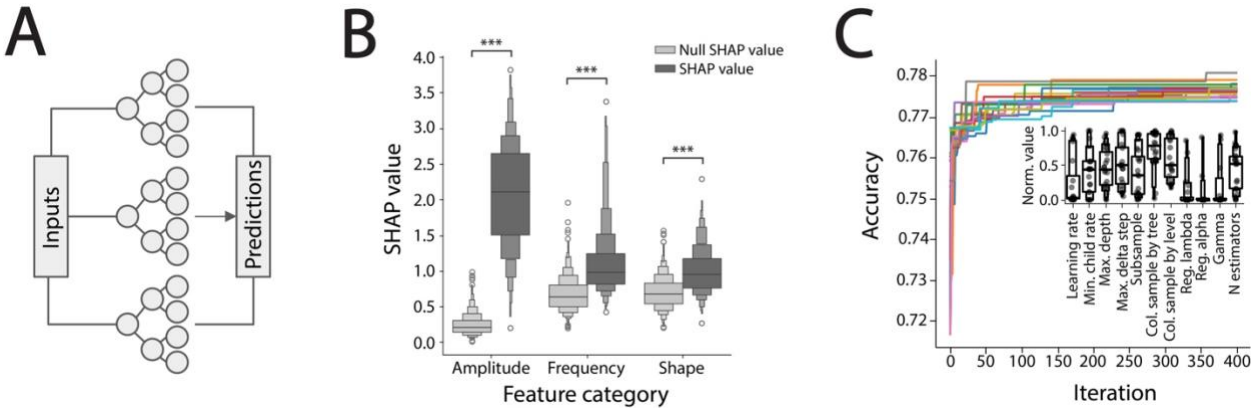

Figure 2. Characterizing the classification model. A) Schematic of a XGBoost model represented as a set of tree-based models. **B)** Interrogation of feature importance using SHAP values shows amplitude to be the most important feature, followed by the sets of frequency and shape-related features and all sets of features show stronger contribution to predictions than expected by change (Null SHAP values; *** $p < 0.001$, Wilcoxon signed-rank). **C)** Classifier model performance is robust to hyperparameter specification. Models trained without hyperparameter optimization exhibit high cross-validation accuracy and hyperparameter optimization results in only small accuracy gains. **Inset)** Distributions of final sets of hyperparameters (normalized to search range) following optimization. All parameters span almost the entire search range but some show tighter clustering.

We next contrasted the performance of the XGB classifier with that of the BSA and QDA classification algorithms. **Figure 3A** shows, for all three classifiers, an example of predicted

occurrences of predicted behavioral labels of each EMG waveform across multiple trials of four tastes of a single recording session. We began our evaluation with gapes, which is the behavior that BSA and QDA were designed to identify using solely frequency characteristics of the EMG signal. Even casual inspection of the panels demonstrates that, as previously shown (Li et al., 2016; Mukherjee et al., 2019; Sadacca et al., 2016), gapes are consistently identified by all classifiers as occurring largely in response to the unpalatable tastant, quinine, and to appear, at earliest, 500ms following stimulus delivery.

To quantify any improvement in performance, we compared the accuracy of each of the different algorithms' classification of user-scored ("ground-truth") gapes—and found that our XGB classifier predicted gape responses marginally better than QDA and significantly better than BSA-based predictions (**Fig. 3B**; top left value in each matrix), despite the latter algorithms being specifically designed for that task.

The remainder of the **Figure 3B** matrix shows how well the classifiers predicted "ground-truth" for other behaviors. Note that QDA is incapable of making MTM predictions, and that while BSA technically does generate a specific "nothing" prediction, this approach relies on the assumption that any orofacial movements with frequencies <4 Hz are insignificant. **Figure 3C** summarizes these data, showing the accuracy weighted to normalize for the very different proportions for behavioral classes in the dataset (i.e., the fact that the majority of identified behaviors were MTMs). Unsurprisingly, QDA performs most poorly, as it cannot account for an entire class (MTMs); The performance of XGB, meanwhile, is better that of either other classifier.

In an attempt to more fairly compare QDA (with its missing MTM label) to BSA and XGB, we moved on to a comparison of all 3 algorithms in a binary classification task, using only the categories QDA is capable of predicting (gape and no movements); all algorithms performed similarly well for binary classification of gapes, but for binary classification of nothing movements, we find XGB once again shows the highest performance, followed by BSA, and QDA performing most poorly (see **Supplementary Figure 10**, also see below for D' results on binary classification).

Of course, the above measures do not provide a full evaluation of classifier performance, because a truly competent model needs to have both high precision (the fraction of predicted hits that are "true" hits) and recall (the fraction of "true" hits that are predicted). An analysis of these two metrics is presented in **Figure 3D.** XGB performs on par with QDA with regard to gapes, and both perform slightly better than BSA. For MTMs, meanwhile, both BSA and XGB show high precision, but XGB has significantly higher recall than either i.e., it correctly identifies many more true positives. Finally, for "nothing" movements, all classifiers have similar recall but only XGB shows high precision.

As high recall is of limited use without (at least reasonably) high precision (e.g., a classifier that gives all datapoints the same label will have perfect recall for that class but poor precision overall), the most rigorous evaluation comes from aggregating the two measures into a weighted F1-score. As shown in **Figure 3E**, QDA performs the worst due to its inability to predict the MTM label, BSA does slightly better, and the XGB classifier shows the highest F1-scores. Overall, XGB best balances recall with precision to identify known behaviors with significantly improved accuracy.

As a convergent test, we compared the classifiers using the d' metric (Fisher Discrimination Index, see Methods), which provides a similarly bias-free measure of detection ability. Derived from signal detection theory for binary classification tasks, d' is calculated by transforming the difference between true positive rate and the false positive rate into a Z-score. As a Z-score, d' values between 1 and 2 are considered to indicate weak performance while those higher than 2 indicate strong performance. Given this rubric, we find that for gapes, both QDA and XGB show strong performance while BSA performs weakly. For MTMs, meanwhile, BSA again performs weakly (likely due to its poor recall) while XGB performs strongly. Finally, for the "nothing" category while both QDA (2.27±0.89) and BSA (2.25±1.25) perform strongly, XGB significantly outperforms both (3.82±1.02, mean±SD; **Fig. 3F**). In summary, XGB is the only algorithm for which d' > 2 for all classes.

We performed one last independent test of the prediction quality of the XGB classifier, assessing the specificity of prediction by determining the fraction of taste-responsive movements (gapes and MTMs) that occurred post-stimulus delivery. Logically, the animal should be making almost exclusively "nothing" movements in the absence of a stimulus—that is, gapes and MTMs should largely be confined to the post-stimulus time-period. Our test of this "post-taste specificity", which is quantified in **Figure 3G** shows that the XGB classifier had significantly higher specificity than BSA for the prediction of gapes, and performed on par with QDA (QDA vs BSA $t(8) = 6$, $p < 0.001$, XGB vs BSA $t(8) = 6.6$, $p < 0.001$, QDA vs XGB $t(8) = 1.2$, $p = .277$, paired t-tests for all comparisons), while XGB matched BSA for specificity of MTMs ($t(8) = 1.3$, $p = 0.228$, paired t-test). Again, QDA is unable to predict MTMs altogether.

The above evidence makes it abundantly clear that the XGB classifier outperforms previous algorithms even on tasks that the previous algorithms were specialized to handle—a fact that takes on even greater significance in light of the fact that the XGB classifier also has the capacity to (verifiably) classify categories of behaviors that previous algorithms cannot handle. This validation gave us the confidence to apply XGB to 22 additional test sessions with AD EMG recordings (n = 9 animals; see Methods), and to use model predictions to characterize the structural and temporal properties of MTMs.

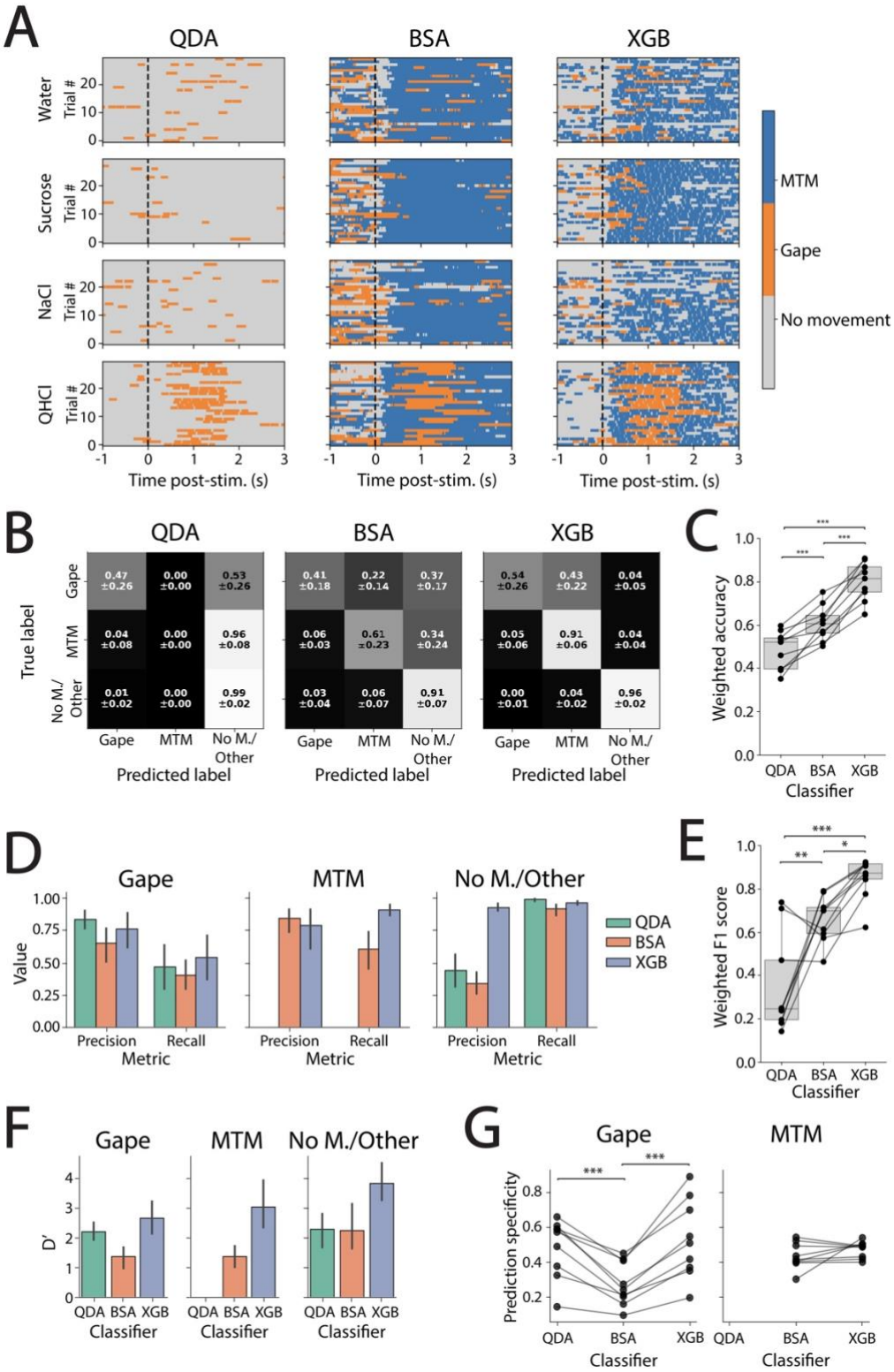

Figure 3. Comparison of XGB with previously used classifiers. A) Example raster plot of predictions for QDA, BSA, and XGB classifiers for four tastants. **B)** Confusion matrices showing class-specific predictions for all three models. **C)** Weighted accuracy for all three models. XGB outperforms both QDA and BSA (*** $p < 0.001$, Tukey post-hoc). **D)** Breakdown

of prediction precision and recall for all three models across label class. XGB performs on par with or better than QDA and BSA for all classes on both metrics. **E)** Weighted F1-score for all 3 models. XGB again outperforms both QDA and BSA (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Tukey post-hoc). **F)** D' values for binary classification by all 3 algorithms on all classes. XGB consistently performs best for all class labels. **G)** Comparison of temporal specificity of rejection/ingestion-related responses (*** $p < 0.001$, paired t-tests).

MTMs are heterogenous and comprised of three distinct behavioral subtypes

The primary goal of the current study is to explore whether MTMs are complex and non-uniform, and test the hypothesis that the decision to consume is characterized by a change in MTM structure. Since the MTM class is defined here as an amalgam of two behaviors (see Methods), it is reasonable to expect that they will not manifest as a single, uniform type when examining EMG features. Although with visual inspection alone (of either video or EMG data) it is difficult to classify or label any such subtypes, we hypothesized that our computational analyses of EMG features might enable their detection. Any such subtypes would be expected to emerge as distinct clusters in feature space. If such clusters are found, we can then attempt to align them across sessions, providing a rigorous test of the clusters' validity (and improving statistical power).

In preparation for these analyses, we linearly reduced the dimensionality of the feature data using Principal Component Analysis (PCA), a step that has been shown to improve model fitting quality (Aggarwal et al., 2014) while preserving most of the data's original structure. All clustering analyses were thereafter conducted in 6-dimensional PCA space (reduced from 8 dimensions, retaining 90% of the variance). UMAP (Uniform Manifold Approximation and Projection, a stochastic, non-linear transform that distorts distances) was used only for visualization since it represents the data qualitatively.

To determine the best number of clusters to describe the data, we iteratively applied Gaussian Mixture Modeling (GMM) with 1-14 clusters to the PCA-reduced waveform features in each individual test session. The number of clusters that best describe the data was identified using Bayesian Information Criterion (BIC), a goodness-of-fit metric that accounts for model complexity to prevent overfitting. **Figure 4A** shows a representative example of this analysis, which reveals that MTM features, shown in UMAP space, are best described as consisting of three distinct clusters of behavior subtypes.

This description of MTMs as being made up of three specific clusters was robust across multiple fits to the same dataset (50 repeats for each test session), multiple test sessions, and multiple rats (**Fig. 4B**; mean number of clusters at lowest BIC value = 3.3). We therefore simplified subsequent analyses by using models specifying three clusters for every test session. As a validation step, we tested whether the clusters were truly distinct within sessions by comparing pairwise Mahalanobis distances between and within clusters; this test confirmed that inter-cluster distances were significantly larger than intra-cluster distances (D statistic = 0.94, $p < 0.001$, Kolmogorov-Smirnov test; **Fig. 4C**). Together, these analyses suggest that MTMs are a category umbrella over three subtypes of behavior—i.e., MTMs are not a simple amalgam of tongue protrusions and mouth movements.

To facilitate across-session comparisons of these results, we aligned cluster indices, which are randomly assigned during the initial clustering process (for more on the "label-switching problem" in mixture models, see Stephens, 2000). Using a "greedy approach", we calculated the cosine similarity between all clusters from two test sessions and identified the index alignment that yielded the highest similarity. The aligned clusters were then merged and compared to the next test session (schematized in **Fig. 4D**) until all sessions were aligned. To visualize all waveforms across all sessions, we projected each waveform into UMAP space (**Fig. 4E**). In this representation, it is apparent that the clusters occupy overlapping yet distinguishable regions with clearly separated density peaks (red dots). To quantitatively check the distinctiveness and stereotypy of clusters across sessions, we trained a support vector machine classifier on cluster features using a leave-one-animal-out cross-validation scheme. This classifier predicted cluster index significantly above chance ($t(32) = 10.36$, $p < 0.001$, one sample t-test; see **Fig. 4F**), confirming that the three clusters are consistently distinct across animals.

The distinguishability of clusters implies that waveforms in one cluster differ significantly from those in the others. To directly test this implication, we first visualized the archetypal shape of waveforms in each cluster by plotting the waveforms closest to each cluster's peak density. When viewed side by side, these waveforms exhibit clear differences in shape and appear to parametrically vary with regard to symmetry (**Fig. 4G**).

To quantitively assess differences in features between the clusters, we performed Kruskal–Wallis tests (the lack of normality in the data distributions precluded the use of parametric statistics), which revealed significant differences between clusters for all features (all $p < 0.001$). Note, however, that effect sizes varied from negligible or weak for most features ($\varepsilon^2 < 0.04$), to moderate for Principal Component (PC) 2 and time to previous peak ($\varepsilon^2 = 0.1$ and 0.04, respectively), to relatively strong for PC 1 ($\varepsilon^2 = 0.23$; see **Table 3**). As PC 1 had the greatest effect size, showed significant differences across all clusters ($p < 0.001$ for all pairwise comparisons, Mann-Whitney U-Test; **Fig. 4H**), and is derived directly from waveform shape, this suggests that the overall EMG waveform shape does a good job of distinguishing the different MTM clusters. Hence, to dissect how the shapes differed, we measured an additional set of descriptive features derived from each waveform's shape: area under the curve, rising slope, falling slope, full width half max, symmetry, and skewness. Kruskal–Wallis tests revealed significant differences between clusters for all metrics ($p < 0.001$ for all comparisons), with symmetry of the waveforms having the greatest effect size ($\varepsilon^2 = 0.12$; see **Table 3**). This quantification confirms the differences in symmetry visually apparent in the representative waveforms in **Figure 4G**.

We conclude, therefore, that MTMs are comprised of three distinct behavioral subtypes which are consistent across sessions and rats. The EMG waveforms across clusters differ by shape, and most prominently by symmetry. Although we cannot infer directly how these waveforms reflect visual differences in orofacial movements (again, these MTMs were difficult to distinguish by eye), the results suggest meaningful differences in the temporal dynamics of the muscle contractions—differences that have been shown using biophysical modelling to impact the shape of the movement, particularly in the context of other muscle activity (see Discussion, and Ross et al., 2024).

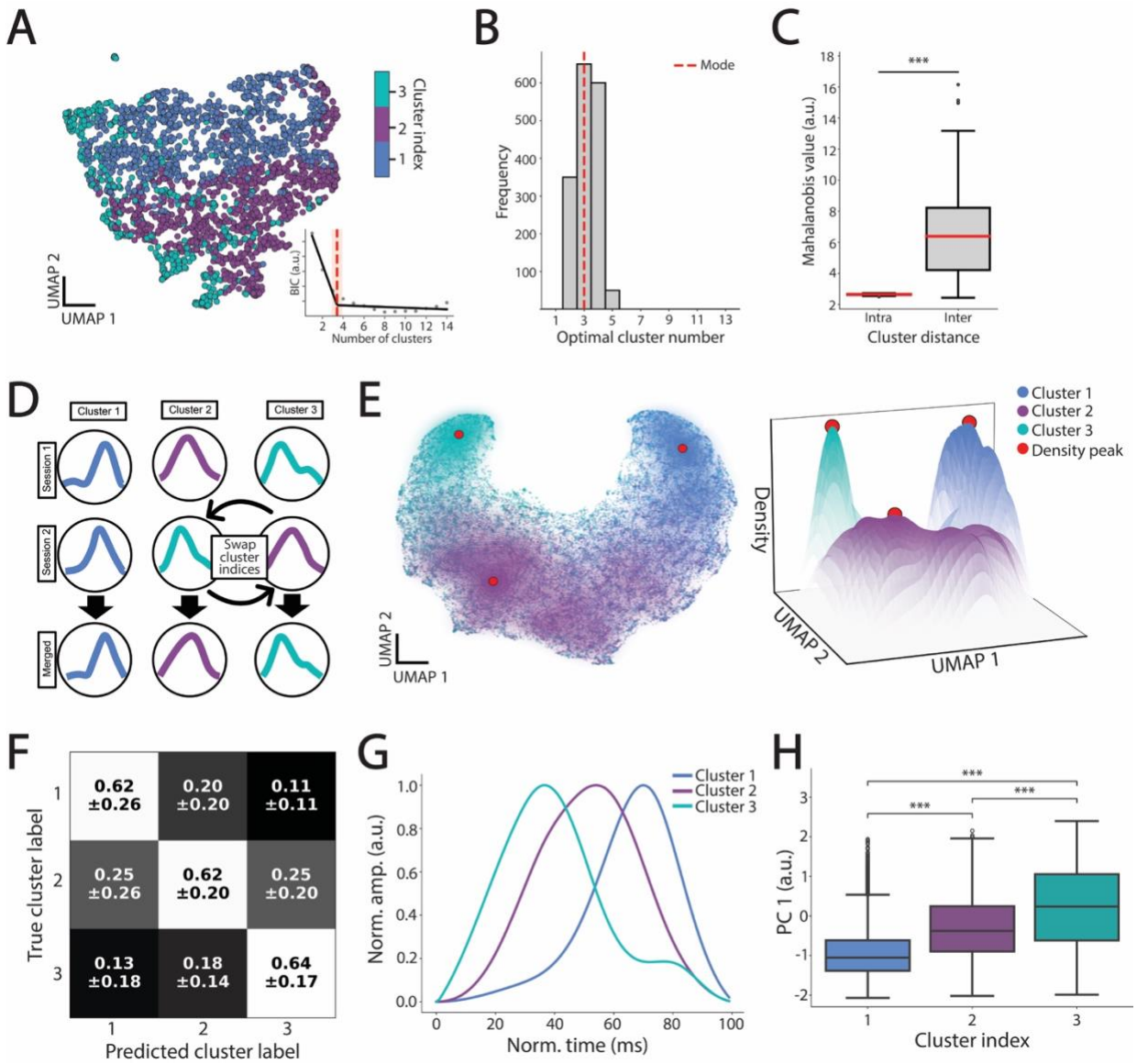

Figure 4. MTMs are comprised of three distinct behavioral subtypes. A) Example MTM waveforms in UMAP space, clustered into three groups. **Inset)** BIC values for 1-14 clusters indicating that three clusters are optimal for this dataset, based on the elbow method. **B)** Distribution of optimal cluster number across 50 iterations for each test session (histogram mode and mean = 3). **C)** Comparison of within clusters (intra-cluster) versus between clusters (inter-cluster) Mahalanobis distances indicates high separability of clusters for each session (*** $p < 0.001$, Kolmogorov-Smirnov test). **D)** Schematic depicting cluster index alignment across test sessions using cosine similarity. **E)** MTM waveforms across all test sessions projected into UMAP space, colored by cluster index label. (Left) Two-dimensional view highlighting the spread of the MTM distribution. (Right) Surface plot showing cluster densities with density peaks (red dots) clearly separable in space. **F)** Confusion matrix from a SVM (support vector machine) classifier showing robust cluster prediction accuracy for leave-one-animal-out cross-validation. **G)** Representative waveforms for each cluster, selected as those closest to the density peak. **H)** The first principal component (PC 1) values, which are derived from waveform shape, are significantly different across clusters (*** $p < 0.001$, Mann-Whitney U-Test).

Table 3. Summary of Kruskal-Wallis test results for feature differences between clusters.

Feature	H	p	ε^2	Effect size
Duration	1339.24	< 0.001	0.0325	Weak
Distance from last peak	1842.74	< 0.001	0.0448	Moderate
Distance from next peak	629.23	< 0.001	0.0153	Weak
Maximum frequency	1112.94	< 0.001	0.0270	Weak
Normalized amplitude	43.48	< 0.001	0.0010	Negligible
Principle component 1	9587.23	< 0.001	0.2333	Relatively strong
Principle component 2	4132.21	< 0.001	0.1005	Moderate
Principle component 3	194.32	< 0.001	0.0047	Negligible
Amplitude	29.67	< 0.001	0.0006	Negligible
Area under the curve	8.71	< 0.001	0.0001	Negligible
Full width half max	206.71	< 0.001	0.0049	Negligible
Rising slope	281.80	< 0.001	0.0068	Negligible
Falling slope	552.79	< 0.001	0.0134	Weak
Symmetry	4804.77	< 0.001	0.1169	Moderate
Skew	1464.57	< 0.001	0.0356	Weak

Frequency of MTM subtypes change around the expected time of the consumption decision

The fact that MTMs are comprised of multiple behavioral subtypes begs the question of whether these MTM clusters occur uniformly across the taste response, or whether their frequencies change when the rat decides to ingest. We started with the hypothesis that this change would occur at around 800 ms post-taste delivery, the approximate time of the previously described transition of GC taste responses into palatability-relatedness (Katz et al., 2001; Mukherjee et al., 2019; Sadacca et al., 2012, 2016). Our initial look at this question is a general one, in which we ask whether the organization of MTMs as a group change across trial halves. If such a shift occurs, we can then examine two competing possibilities: 1) a specific subtype of MTMs is associated with the onset of ingestion, or 2) there are changes across MTMs in general around the consumption decision.

To perform this initial test, we divided MTMs into two categories on the basis of whether they occurred “before” or “after” 800 ms post-taste delivery. For each test session, we then reduced the dimensionality of MTM features in each category using the UMAP algorithm

(Uniform Manifold Approximation and Projection) to visualize MTM feature distributions. This process revealed distinct shifts between the "before" and "after" categories in the distributions of MTM features (**Fig. 5A**). We quantified whether these differences were significant by subtracting "before" from "after" binned density histograms and summing the absolute values of non-zero bins. Statistical significance was assessed with a permutation test: 1,000 bootstrap shuffles of MTM events were used to generate a null distribution and p-values were calculated as the proportion exceeding the observed divergence. As UMAP is a stochastic embedding (meaning the reduced data does not look the same if the transformation is repeated), we repeated the transformation and assessment of the differences 10 times to ensure robustness of our results. The median p-value for all sessions was below 0.05, signifying that the MTMs produced before the approximate time of palatability-related firing are distinguishable from those produced after.

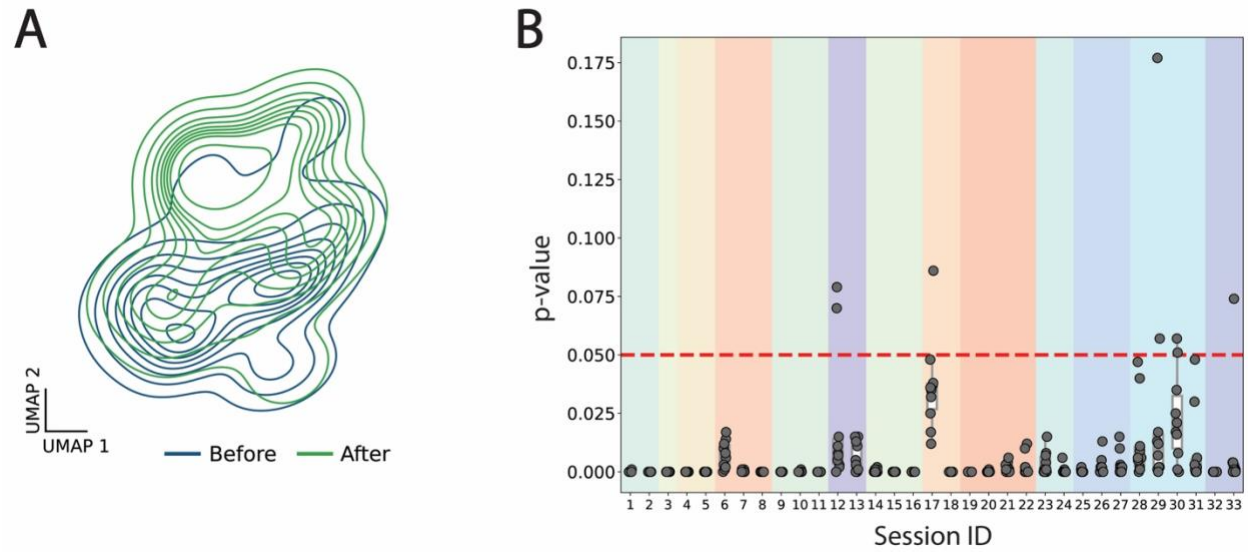

Figure 5. MTM features shift across the taste response. A) Example contour plot of UMAP-reduced MTM feature distributions categorized by their timing relative to average palatability activity onset ("before" [blue] versus "after" [green] 800 ms post-stimulus delivery). **B)** Session-wise p-values quantifying the shift in MTM feature distributions before versus after average palatability activity, calculated using bootstrap resampling. The median p-value across multiple repeats for all sessions is <0.05. Colors denote sessions from the same animal.

We moved on to testing whether a single MTM cluster might be the ingestive homologue to gapes—i.e., similarly occurring mainly after GC's transition to palatability-related activity (Mukherjee et al., 2019; Sadacca et al., 2016)—or whether we would observe frequency changes across multiple MTM clusters, suggestive of a more subtle behavioral pattern shift. We analyzed the three MTM clusters and gapes within a 1 sec window centered at 800 ms post-stimulus delivery, corresponding to the average palatability transition time (Sadacca et al., 2012, 2016). We specifically hypothesized that if palatable tastants elicit broad ingestive-related behavioral changes, this should be reflected as significant changes in the frequencies of occurrence multiple MTM clusters.

Figure 6A illustrates, for an example session, the occurrence of MTM cluster #2 (MTM2) across 30 sucrose trials. The occurrence of this particular MTM cluster increases

dramatically after the average transition time. But further analysis made it clear that MTM2 was not the only behavior to change with the consumption decision. We counted the number of each category of behavior on each trial occurring before and after the 800 ms timepoint, and, for each tastant within each session, tested whether the frequency of each movement significantly changed after the transition. In the example distributions presented in **Figure 6B**, gapes significantly increased during quinine trials (stat = -2.06, $p = 0.04$, Poisson means test), as expected. Meanwhile, during sucrose trials, gapes, MTM1, and MTM2 all significantly changed (all $p < 0.05$, Poisson means test).

We next assessed whether these broad changes were reflected across all sessions, tastants, and behaviors. To avoid issues of multiple comparisons, we calculated the number of significant changes collectively (473 comparisons) and tested whether this number exceeded the proportion expected by chance, by comparing against the distribution of p-values expected under the null hypothesis (**Fig. 6C**). A binomial test confirmed that the number of observed significant p-values were greater than expected by chance ($p < 0.001$). These analyses indicate the presence of significant changes in the frequencies of behavior across time, however, it does not tell us how these changes are distributed across tastes and behaviors.

We hypothesized that these shifts are driven primarily by gapes in response to aversive tastants, whereas palatable tastants elicit patterns driven primarily by MTMs. We further predicted that the consistency of the pattern change should scale with the tastant's valence. A tastant-by-behavior significance matrix, showing the fraction of significant test sessions for each tastant-behavior combination (**Fig. 6D**), confirmed these hypotheses. All tastant-behavior combinations were significant above chance (that is a fraction >0.05 of tests were significant, with a single exception of MTM3 to citric acid), highlighting the robustness of the change. Examining the aversive tastants, gapes are most consistently changing to the highly aversive quinine (QHCl), recapitulating previous findings (Mukherjee et al., 2019; Sadacca et al., 2016). Dilute quinine (Dil-QHCl) showed weakly consistent changes across all behaviors, which is unsurprising given reports that this low concentration elicits very few gapes (Grill & Norgren, 1978a). In addition, citric acid unexpectedly elicited a highly consistent change to MTM1 frequency, a result we address in the Discussion.

For palatable tastants, we observed highly consistent changes across all three MTM clusters. For MTM1 and MTM2, the consistency of these changes scaled with palatability: sucrose elicited the most consistent changes (both MTM1 and MTM2 showed high fractions of significant comparisons, 0.42 and 0.61 respectively), followed by sodium chloride [NaCl], then water (which only showed relatively high fractions from MTMs, 0.44 and 0.38 for each taste respectively). Altogether these results suggest that the GC palatability transition elicits reliable but broad changes in behavior as part of ingestion, extending our understanding beyond the previously demonstrated onset of gaping to quinine.

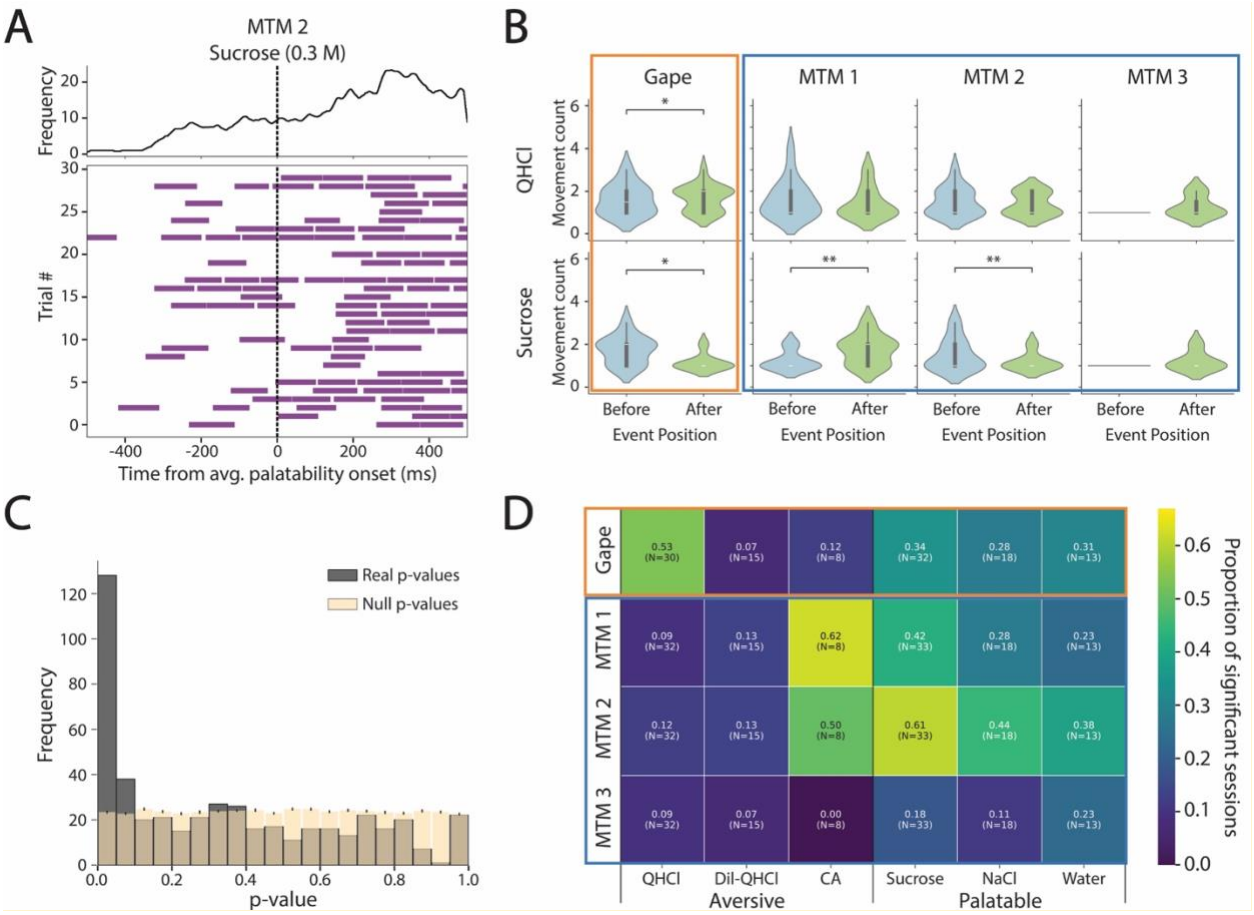

Figure 6. Frequency of all MTM subtypes changes across mean consumption response time. A) Raster plot of occurrence of MTM cluster #3 across 30 trials of sucrose in an example test session, aligned to the average palatability activity onset (800 ms post-stimulus delivery). **B)** Counts of the number of occurrences of gapes (left column) and three MTM clusters (right columns), comparing the number of behavioral events that occurred 500 ms before versus after palatability onset for quinine (top row) and sucrose (bottom row) trials in an example session (* $p < 0.05$, ** $p < 0.01$, Poisson means test). **C)** Distribution of Poisson means test p-values across all tastant-behavior comparisons for every test session (gray) compared to a null distribution (yellow). **D)** Tastant-behavior matrix, with brightness of color indicating the proportion of significant test sessions showing broad changes in behavior frequency.

MTMs change with Gustatory Cortex's transition to palatability-related firing

The above results support our most risky hypothesis, in that they reveal broad changes in the occurrence probability of MTM clusters elicited by palatable tastants, centered on the average time at which GC taste responses have been previously shown to transition into a period of palatability-related firing which drives gaping onset (Katz et al., 2001; Mukherjee et al., 2019; Sadacca et al., 2012, 2016). To more directly test this link in an independent manner that does not presuppose a specific time point for a behavioral change, we analyzed sessions in which EMG and GC single-neuron taste responses were simultaneously recorded (n = 22 test sessions, N = 9 animals; see Methods), characterizing the timing of nonlinear changes in both ensembles

of neurons and ensembles of behaviors. Specifically, we fit changepoint models to time series of behaviors to determine when significant changes in the occurrence probabilities of the behavior occur (again, this analysis allows us to evaluate all MTM subtypes collectively, rather than individual types in isolation) and compared the timing of these inferred changes to onsets of palatability-related ensemble activity in GC.

We began by determining the number of states that best describe the dynamics of the classified orofacial behaviors. We iteratively fit change point models with different numbers (2-8) of states to 2000 ms of post-stimulus activity, and compared the models using the ELBO (**E**vidence **L**ower **Bo**und, not to be confused with the "elbow method" used to evaluate BIC in the previous section), a metric optimized to fit Bayesian models using variational inference (see Methods). Models with 3 states (that is, 2 changepoints) provided by far the best fit to the data (**Fig. 7A**). Follow-up testing showed that the 3-state solution had significantly lower ELBO than models with any other number of states ($p < 0.005$ for all pairwise comparisons, t-test with Bonferroni correction), and hence the 3-state model was used for all subsequent analyses. **Figure 7B** shows a representative example of data from a single session for behaviors observed across all tastes (left column), the timing of the states and changepoints inferred by the model (middle column), and the probabilities of each event occurring in a particular state for any given taste (right column).

Furthermore, a 2-way ANOVA (states x tastes) on ELBO values for all sessions showed significant differences between the number of states ($SS = 172.9$, $DF = 6$, $p < 0.001$) but not between tastes ($SS = 0.005$, $DF = 4$, $p = 0.99$), and we found no significant differences between tastes for the ELBO in 3-state models ($p > 0.1$ for all pairwise comparisons between tastes, t-test with Bonferroni correction). These results make it clear that the 3-state solution describes the behavioral results of taste processing in general (rather than being the response to a particular taste or subset of tastes). Given that fact, we pooled model results across tastes for subsequent analyses and found that the median of the distribution of first changepoint latencies occurred early in the latter half of the 1st second post-stimulus delivery, and the median of the second changepoint more than half a second later.

Next, we brought an analogous analysis to bear on the neural ensemble data, measuring the evolution of palatability encoding in the recorded GC population ($n = 143$ neurons, 7 sessions, 27.3±-4.1 neurons/session, mean±SD), and distilling the population's palatability response down to a single time series using PCA (see Sadacca et al. 2012). We then fit a sigmoid to this first principal component (PC 1, blue crosses in **Fig. 7C**); the inflection point of this sigmoid was used to infer when the change in palatability encoding occurs (black line, **Fig. 7B**).

This analysis allowed us to perform a comparison of brain and behavior that strongly supported our central hypothesis, in that the median occurrence of the transition to palatability-related firing occurs at almost precisely the time of the first behavioral transition. **Figure 7D** shows the posterior distribution (i.e. inferred uncertainty) in the estimate of the inflection point of this sigmoid. We compared this inflection point distribution to the distribution of MTM changepoints pooled across sessions and tastes (to maximize statistical power, **Fig. 7E**), and found that the latency of palatability-response onset was significantly earlier than that of MTM changepoint 2 ($stat = 1214$, $p < 0.001$, Mann-Whitney U-Test), but not different from the latency distribution of MTM changepoint 1 ($stat = 4021$, $p = 0.72$, Mann-Whitney U-Test). That is, changes in the occurrence of the ensemble of behaviors are aligned with the specific time point at which GC neural ensemble responses become palatability-related. We performed a Monte-Carlo

simulation to test for the possibility that this alignment of the palatability inflection could have happened by chance with either of the average MTM changepoint locations (see Methods, and Stone et al., 2022 for a similar analysis) and found this to be unlikely ($p = 0.028$).

Finally, to test whether the behavior change points in responses to palatable tastes were truly driven primarily by MTMs and not gapes, we performed a pair of analyses. First, we demonstrated that the specific probabilities of the MTM clusters change with state (SS = 1.21, DF = 2, $p < 0.001$), in line with results in the previous section. As a convergent analysis, we assessed the extent to which removing gapes and "nothing" movements reduced changes between subsequent states inferred by the changepoint model (**Fig. 7F**). As expected, the results of this analysis were taste-specific (2-way ANOVA, category [all vs. MTM only]: SS = 0.15, DF = 1, $p = 0.52$, category*taste: SS = 10.4, DF = 4, $p < 0.001$): the difference between subsequent states for the "MTM only" (gapes and nothing removed) condition was significantly decreased compared to the "all behaviors" condition for quinine; that is, pre- and post-decision states became more similar when gapes were removed from quinine trials ($p < 0.001$, t-test with Bonferroni correction) but gape and "nothing" removal had no impact on the behavior change for any other taste ($p > 0.1$, t-test with Bonferroni correction).

These results demonstrate that gapes most strongly affect the decision-related behavior change in response to quinine. In contrast, MTMs (in a distributed manner across all subtypes) are primarily driving the behavior changes in response to palatable tastant (see **Supplementary Figure 11** for a full breakdown of each behavior across states and tastes). We further conclude that this behavioral evidence of an ingestion decision, as detected by the changepoint model, is associated with the same neural transition in GC taste responses that has previously been shown to be responsible for the onset of gaping behavior.

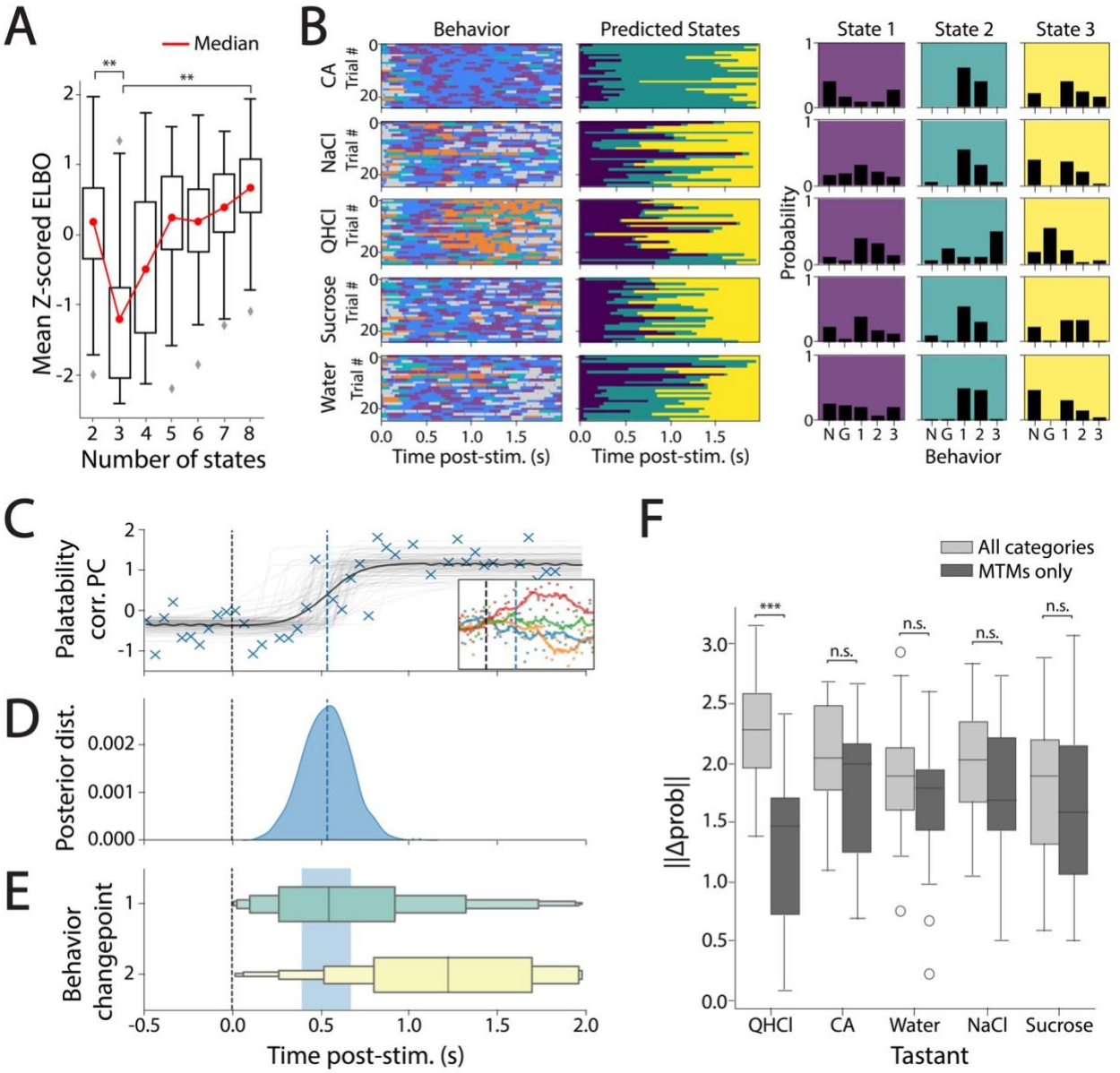

Figure 7. Emergence of palatability encoding in GC neural activity coincides with changepoints in MTM occurrence. A) Comparison of changepoint models with 2-8 states applied to timeseries of oral behaviors reveals that 3-state models best describe behavior dynamics in 2000 ms post-stimulus delivery (** $p < 0.01$, t-test with Bonferroni correction). **B)** Representative example of inference performed by changepoint model. Left column = Raster plot of behaviors across tastes, Orange = gapes, Gray = no movement, Blue = MTM1, Purple = MTM2, Turquoise = MTM3; Middle column = State sequences inferred by changepoint model; Right column = Histograms of probabilities for each behavior per state and taste inferred by changepoint model. N = no movement, G = gape, 1, 2, 3 = MTM cluster 1, 2, and 3, respectively. **C)** Sigmoid fit to first principal component of pooled palatability correlation timeseries ("Palatability corr. PC"). **Inset)** Representative example showing projection of activity of a single neural population (session) along the palatability axis. Black line = Stimulus delivery, Blue line = Sigmoid inflection point, Red = Sucrose, green = NaCl, Blue = Quinine, Yellow =

Citric Acid. **D)** Posterior distribution for inflection point of sigmoid fit. **E)** Distributions of each changepoint in the 3-state model pooled across tastes and sessions. Shaded blue square denotes mean±SD of the distribution in panel D. **F)** Comparison of normalized magnitude of difference between inferred behavior probability vectors for subsequent states, including either all categories or MTMs only. Only quinine (QHCl) shows significant decrement in difference of state-specific behavior probabilities (states are more similar) when gapes and no movements are removed (*** $p < 0.001$, t-test with Bonferroni correction).

Discussion

Detection of orofacial movements using machine learning on EMG activity

A stimulus on the tongue necessitates a binary decision: to ingest or reject. The fact that the choice is binary does not necessarily mean that it is simple, however. Grill and Norgren (1978a) proposed that rats put this decision into action through highly stereotyped behaviors such as gapes, the primary rejective orofacial behavior, which are easily detected both visually and *via* EMG activity of the anterior digastric (AD; jaw opener) muscle. This has led to multiple investigations into how the taste system drives gaping (Dinardo & Travers, 1994; DiNardo & Travers, 1997; Kinzeler & Travers, 2008; Li et al., 2016; Mukherjee et al., 2019; J. Travers et al., 2000). Meanwhile, similar studies probing ingestive behaviors are few, due in large part to the complexity of and lack of consensus regarding the two primary ingestive-related behaviors: mouth movements and tongue protrusions (Grill & Norgren, 1978a). In this study, we have leveraged recent advances in computational analysis in a reexamination of the ingestive response, in the process investigating how the evolution of behavior is linked to activity in Gustatory Cortex (GC).

To perform these analyses, we first trained a machine learning classifier to detect and classify individual orofacial movements using AD electromyography (EMG) activity. We targeted the AD specifically because its contraction is required for the execution of every movement cycle and differs significantly in duration between gaping and licking (J. B. Travers & Norgren, 1986), and the muscle is readily accessible for EMG implantation. We chose to use EMG activity not only because it allowed us to collect these data in freely moving rats but, more importantly, because it allowed us to avoid the inherent difficulties in distinguishing mouth openings from tongue protrusions in video recordings (in which the appearance of these behaviors is similar). Instead, to perform data labelling, we defined an umbrella category of rapid mandible openings and closings with or without tongue extension—called mouth or tongue movements (MTMs)—that could then be parsed into finer subtypes later (see below) based on features in the EMG signals. In addition, EMG provided several further practical advantages over traditional video-based scoring, including higher temporal resolution, increased data throughput (e.g., no trial loss from camera obstruction), computational efficiency, and reduced bias from scoring by an experimenter.

The XGBoost classifier (XGB) presented in this paper performed robustly across multiple metrics, outperforming previous algorithms even when validated on detection of gapes, the behavior that these previously published classifiers were specifically designed to detect (Li et al., 2016; Mukherjee et al., 2019). This performance improvement almost certainly reflects XGB's use of a suite of features—notably amplitude, frequency, and waveform shape—that together enhance discrimination of EMG waveforms generated by different behaviors.

Detection of three mouth or tongue movements (MTMs) subtypes

The suggestion that there is a distribution underlying MTMs is not unlike to the model proposed by Kaplan and Grill (1995) whereby mouth movements and tongue protrusions were suggested to exist on a continuum of tongue extension (delineated by the tongue crossing the incisor plane). In fact, these authors specifically proposed that a more advanced analysis of EMG activity was needed to elucidate this continuum. Here, we have provided precisely this analysis,

which reveals that MTMs consistently group into not two but three stereotyped clusters that are robust across sessions and animals. While the clusters were significantly differentiable, overlap between them suggests that MTM subtypes are not perfectly distinct categories, but rather best described as parts of a single tri-modal distribution of movement patterns. The trimodal distribution of MTMs revealed in our EMG analysis may help explain the ambiguity in the literature regarding the visual distinction between mouth movements and tongue protrusions (Berridge, 1991; Berridge & Grill, 1983; Berridge & Treit, 1986; Brining et al., 1991; De Luca et al., 2012; Feurté et al., 2000; Fontanini & Katz, 2006; Jarrett et al., 2005; Kaplan et al., 1995, 1995; López et al., 2023; Parker et al., 1992; Pereira et al., 2021; Soares et al., 2007; see Introduction).

To better understand how waveforms identified as belonging to separate clusters differed from each other, we produced a set of descriptive features. Waveform shape parameters, and most notably symmetry, best differentiated clusters, suggesting differences in temporal dynamics of the muscle contractions. While it remains difficult to speculate on how this translates into specific mouth shapes (because all MTMs look the same in our video records and only become detectably different with the high temporal resolution of EMG), this may be a useful avenue of investigation *via* video-based data mining in a restrained paradigm (Bollu et al., 2020; Inaba et al., 2025) or biomechanical modelling in future work. Regardless, it is likely that these subtypes represent different modes of intraoral fluid transport (Ross et al., 2024).

The relevance of GC activity to ingestive decisions

Having established that there is an ensemble of three distinct MTM subtypes, we went on to show that there were broad significant changes in the ensemble response to all tastants between the early (0-800 ms) and late (800-1500 ms) post-delivery periods. In agreement with previously published findings (Li et al., 2016; Mukherjee et al., 2019; Sadacca et al., 2016), quinine trials elicited highly consistent changes in gape frequency. Meanwhile, palatable tastants elicited highly consistent changes across all three MTM clusters. Notably, for MTM1 and MTM2 (and MTM3 to a lesser extent), the consistency of the behavioral transition scaled with palatability. The fact that a strongly valanced tastant elicits a more reliable and robust behavioral response accords well with previous results (Grill & Norgren, 1978a; Spector et al., 1998; J. B. Travers & Norgren, 1986). Overall, this confirms that, while rejection decisions are actuated by gaping, the decision to ingest is demarcated by a shift in the behavioral pattern across the ensemble of MTMs.

Unexpectedly, we also found that citric acid elicited uniquely consistent within-trial changes to MTM1 frequency at the time of consumption decisions. Although it is canonically considered an aversive taste—and thus expected to elicit gapes—there are reports of acids eliciting mouth movements and/or tongue protrusions in rats and hamsters (Brining et al., 1991; Sasamoto et al., 2001; S. P. Travers, 2002). For reasons that remain unclear, sour tastants may elicit a rejection behavior pattern distinct from that elicited by quinine that consists of a mixture of ingestion and rejection-related behaviors; however, a more detailed investigation rigorously testing this hypothesis is outside of the scope of this study.

Finally, we directly tested whether the transition to palatability-related activity in GC might be modulating ingestion-related behaviors, just as it does gaping-driven rejection (Mukherjee et al., 2019; Sadacca et al., 2016). Application of changepoint modelling to EMG

activity revealed two time points in the first two seconds of the taste response at which there is a rapid shift in the behavior pattern of MTMs. This outcome contrasts with that of analogous modelling on neural data, which reliably reveals 4-5 states during the same 2000 ms of post-stimulus activity (Mahmood et al., 2023; Sadacca et al., 2016). This confirms previous conclusions that only the neural ensemble transition to palatability-related firing is linked directly to behavior. While this is not causative evidence, these data are highly suggestive that GC dynamics drive ingestive behaviors just as with rejection, and strongly mimic results from the lab for the relation between GC activity and gaping (Sadacca et al., 2016), which later interrogation proved to be causally linked (Mukherjee et al., 2019).

Our data indicate that there is no single ingestion behavior directly analogous to gapes. Instead, ingestion involves coordination among a set of three MTM subtypes. We speculate that upon taste delivery, the rat may initiate a "sampling" phase characterized by one MTM pattern followed by an "ingestion" phase characterized by another (see Ross et al., 2024). It is only by considering the ingestive response as a pattern of multiple behaviors, through analyses such as changepoint modelling, that we were able to elucidate how the response evolves in relation to GC activity. Future work could refine behavioral categorization by recording simultaneously from multiple muscles, with the extrinsic tongue muscles being an obvious first candidate (J. B. Travers & Norgren, 1986). Such an investigation would be able to definitively address whether the different subtypes of MTMs we observe form part of different sequences of muscle movements. The use of tools such as seqNMF (Mackevicius et al., 2019) could allow future investigations to detect sequential motifs in the behavior time series to characterize the patterns of MTM subtypes occurring within the states of the ingest response.

Beyond clarifying the ingestive response and its constituent behaviors, our results advance our understanding of the function of GC. We claim that GC is not simply a sensory region, instead its sensory processing is inextricably linked to the generation of a palatability judgment which ultimately influences motor output. We hypothesize that the most obvious candidate for this modulatory signal is the central pattern generator in the medullar reticular formation which has been shown to switch between several oromotor rhythmic patterns (Z. Chen et al., 2001; DiNardo & Travers, 1997; Grill & Norgren, 1978b; Moriyama, 1987; J. Travers et al., 2000; Venugopal et al., 2007). In other words, the present results suggest that both ingestive and rejective oral behaviors are both governed by the same underlying neural mechanisms.

References

Aggarwal, Charu C., and Chandan K. Reddy, editors. Data Clustering: Algorithms and Applications. CRC Press, 2014.

Akaike, H. (1998). Information Theory and an Extension of the Maximum Likelihood Principle. In *Selected Papers of Hirotugu Akaike* (pp. 199–213). Springer, New York, NY. https://doi.org/10.1007/978-1-4612-1694-0_15

Bales, M. B., Schier, L. A., Blonde, G. D., & Spector, A. C. (2015). Extensive Gustatory Cortex Lesions Significantly Impair Taste Sensitivity to KCl and Quinine but Not to Sucrose in Rats. *PLOS ONE*, *10*(11), e0143419. https://doi.org/10.1371/journal.pone.0143419

Berridge, K. C. (1991). Modulation of taste affect by hunger, caloric satiety, and sensory-specific satiety in the rat. *Appetite*, *16*(2), 103–120. https://doi.org/10.1016/0195-6663(91)90036-r

Berridge, K. C., & Grill, H. J. (1983). Alternating ingestive and aversive consummatory responses suggest a two-dimensional analysis of palatability in rats. *Behavioral Neuroscience*, *97*(4), 563–573. https://doi.org/10.1037//0735-7044.97.4.563

Berridge, K. C., & Treit, D. (1986). Chlordiazepoxide directly enhances positive ingestive reactions in rats. *Pharmacology Biochemistry and Behavior*, *24*(2), 217–221. https://doi.org/10.1016/0091-3057(86)90341-2

Berthoud, H.-R., & Münzberg, H. (2011). The lateral hypothalamus as integrator of metabolic and environmental needs: From electrical self-stimulation to opto-genetics. *Physiology & Behavior*, *104*(1), 29–39. https://doi.org/10.1016/j.physbeh.2011.04.051

Bollu, T., Ito, B., Whitehead, S. C., Kardon, B., Redd, J., Liu, M. H., & Goldberg, J. H. (2020). *How a mouse licks a spout: Cortex-dependent corrections as the tongue reaches for, and misses, targets* (p. 655852). bioRxiv. https://doi.org/10.1101/655852

Brining, S. K., Belecky, T. L., & Smith, D. V. (1991). Taste reactivity in the hamster. *Physiology & Behavior*, *49*(6), 1265–1272. https://doi.org/10.1016/0031-9384(91)90361-Q

Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*. 785–794. https://doi.org/10.1145/2939672.2939785

Chen, Z., Travers, S. P., & Travers, J. B. (2001). Muscimol infusions in the brain stem reticular formation reversibly block ingestion in the awake rat. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *280*(4), R1085–R1094. https://doi.org/10.1152/ajpregu.2001.280.4.R1085

De Luca, M. A., Solinas, M., Bimpisidis, Z., Goldberg, S. R., & Di Chiara, G. (2012). Cannabinoid facilitation of behavioral and biochemical hedonic taste responses. *Neuropharmacology*, *63*(1), 161–168. https://doi.org/10.1016/j.neuropharm.2011.10.018

Dinardo, L. A., & Travers, J. B. (1994). Hypoglossal neural activity during ingestion and rejection in the awake rat. *Journal of Neurophysiology*, *72*(3), 1181–1191. https://doi.org/10.1152/jn.1994.72.3.1181

DiNardo, L. A., & Travers, J. B. (1997). Distribution of Fos-Like Immunoreactivity in the Medullary Reticular Formation of the Rat after Gustatory Elicited Ingestion and Rejection Behaviors. *Journal of Neuroscience*, *17*(10), 3826–3839. https://doi.org/10.1523/JNEUROSCI.17-10-03826.1997

Dziak, J. J., Coffman, D. L., Lanza, S. T., Li, R., & Jermiin, L. S. (2019). Sensitivity and specificity of information criteria. *Briefings in Bioinformatics*, *21*(2), 553–565. https://doi.org/10.1093/bib/bbz016

Feurté, S., Nicolaidis, S., & Berridge, K. C. (2000). Conditioned taste aversion in rats for a threonine-deficient diet: Demonstration by the taste reactivity test. *Physiology & Behavior*, *68*(3), 423–429. https://doi.org/10.1016/S0031-9384(99)00202-4

Fontanini, A., & Katz, D. B. (2006). State-Dependent Modulation of Time-Varying Gustatory Responses. *Journal of Neurophysiology*, *96*(6), 3183–3193. https://doi.org/10.1152/jn.00804.2006

Friard, O., & Gamba, M. (2016). BORIS: A free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*, *7*(11), 1325–1330. https://doi.org/10.1111/2041-210X.12584

Grill, H. J., & Norgren, R. (1978a). The taste reactivity test. I. Mimetic responses to gustatory stimuli in neurologically normal rats. *Brain Research*, *143*(2), 263–279. https://doi.org/10.1016/0006-8993(78)90568-1

Grill, H. J., & Norgren, R. (1978b). The taste reactivity test. II. Mimetic responses to gustatory stimuli in chronic thalamic and chronic decerebrate rats. *Brain Research*, *143*(2), 281–297. https://doi.org/10.1016/0006-8993(78)90569-3

Grossman, S. E., Fontanini, A., Wieskopf, J. S., & Katz, D. B. (2008). Learning-Related Plasticity of Temporal Coding in Simultaneously Recorded Amygdala–Cortical Ensembles. *The Journal of Neuroscience*, *28*(11), 2864–2873. https://doi.org/10.1523/JNEUROSCI.4063-07.2008

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Head, T., Kumar, M., Nahrstaedt, H., Louppe, G., & Shcherbatyi, I. (2024). *Holgern/scikit-optimize* [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.11478072

Inaba, S., Uesaka, N., & Tanaka, D. H. (2025). Automatic quantification of disgust reactions in mice using machine learning. *Scientific Reports*, *15*(1), 17573. https://doi.org/10.1038/s41598-025-01244-3

Jarrett, M. M., Limebeer, C. L., & Parker, L. A. (2005). Effect of Δ9-tetrahydrocannabinol on sucrose palatability as measured by the taste reactivity test. *Physiology & Behavior*, *86*(4), 475–479. https://doi.org/10.1016/j.physbeh.2005.08.033

Jones, L. M., Fontanini, A., Sadacca, B. F., Miller, P., & Katz, D. B. (2007). Natural stimuli evoke dynamic sequences of states in sensory cortical ensembles. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(47), 18772–18777. https://doi.org/10.1073/pnas.0705546104

Kaplan, J. M., Roitman, M. F., & Grill, H. J. (1995). Ingestive taste reactivity as licking behavior. *Neuroscience & Biobehavioral Reviews*, *19*(1), 89–98. https://doi.org/10.1016/0149-7634(94)00023-T

Katz, D. B., Simon, S. A., & Nicolelis, M. A. L. (2001). Dynamic and Multimodal Responses of Gustatory Cortical Neurons in Awake Rats. *Journal of Neuroscience*, *21*(12), 4478–4489. https://doi.org/10.1523/JNEUROSCI.21-12-04478.2001

Khalid, S., Khalil, T., & Nasreen, S. (2014). A survey of feature selection and feature extraction techniques in machine learning. *2014 Science and Information Conference*, 372–378. https://doi.org/10.1109/SAI.2014.6918213

Kinzeler, N. R., & Travers, S. P. (2008). Licking and gaping elicited by microstimulation of the nucleus of the solitary tract. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, *295*(2), R436–R448. https://doi.org/10.1152/ajpregu.00189.2008

Levitan, D., Lin, J.-Y., Wachutka, J., Mukherjee, N., Nelson, S. B., & Katz, D. B. (2019). Single and population coding of taste in the gustatory cortex of awake mice. *Journal of Neurophysiology*, *122*(4), 1342–1356. https://doi.org/10.1152/jn.00357.2019

Li, J. X., Maier, J. X., Reid, E. E., & Katz, D. B. (2016). Sensory Cortical Activity Is Related to the Selection of a Rhythmic Motor Action Pattern. *Journal of Neuroscience*, *36*(20), 5596–5607. https://doi.org/10.1523/JNEUROSCI.3949-15.2016

Lin, J.-Y., Mukherjee, N., Bernstein, M. J., & Katz, D. B. (2021). Perturbation of amygdala-cortical projections reduces ensemble coherence of palatability coding in gustatory cortex. *eLife*, *10*, e65766. https://doi.org/10.7554/eLife.65766

Loeb, G. E., & Gans, C. (1986). *Electromyography for Experimentalists*. University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/E/bo5960315.html

López, M., Dwyer, D. M., Begega, A., Jove, C., & Alcorta, E. (2023). An evaluation of hedonic responses in taste-potentiated odor aversion using the analysis of licking microstructure and orofacial reactivity. *Behavioural Processes*, *213*, 104970. https://doi.org/10.1016/j.beproc.2023.104970

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, *2*(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Mackevicius, E. L., Bahle, A. H., Williams, A. H., Gu, S., Denisenko, N. I., Goldman, M. S., & Fee, M. S. (2019). Unsupervised discovery of temporal sequences in high-dimensional datasets, with applications to neuroscience. *eLife*, *8*, e38471. https://doi.org/10.7554/eLife.38471

Mahmood, A., Mukherjee, N., Stone, B., Raymond, M., Germaine, H., Lin, J.-Y., Mazzio, C., & Katz, D. (2025). *katzlabbrandeis/blech_clust: V1.1.0* (Version 1.1.0) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.15175273

Mahmood, A., Steindler, J., Germaine, H., Miller, P., & Katz, D. B. (2023). Coupled Dynamics of Stimulus-Evoked Gustatory Cortical and Basolateral Amygdalar Activity. *Journal of Neuroscience*, *43*(3), 386–404. https://doi.org/10.1523/JNEUROSCI.1412-22.2022

McInnes, L., Healy, J., Saul, N., & Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software*, *3*(29), 861. https://doi.org/10.21105/joss.00861

Moran, A., & Katz, D. B. (2014). Sensory Cortical Population Dynamics Uniquely Track Behavior across Learning and Extinction. *Journal of Neuroscience*, *34*(4), 1248–1257. https://doi.org/10.1523/JNEUROSCI.3331-13.2014

Moriyama, Y. (1987). Rhythmical jaw movements and lateral ponto-medullary reticular neurons in rats. *Comparative Biochemistry and Physiology Part A: Physiology*, *86*(1), 7–14. https://doi.org/10.1016/0300-9629(87)90268-4

Mukherjee, N., Wachutka, J., & Katz, D. B. (2019). Impact of precisely-timed inhibition of gustatory cortex on taste behavior depends on single-trial ensemble dynamics. *eLife*, *8*, e45968. https://doi.org/10.7554/eLife.45968

Parker, L. A., Maier, S., Rennie, M., & Crebolder, J. (1992). Morphine- and naltrexone-induced modification of palatability: Analysis by the taste reactivity test. *Behavioral Neuroscience*, *106*(6), 999–1010. https://doi.org/10.1037/0735-7044.106.6.999

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *The Journal of Machine Learning Research*, *12*(null), 2825–2830.

Pereira, E. D., Faria, A. M., Andrade-Franzé, G. M. F., Menani, J. V., De Luca, L. A., & Andrade, C. A. F. (2021). Sodium palatability in male spontaneously hypertensive rats. *Hormones and Behavior*, *130*, 104952. https://doi.org/10.1016/j.yhbeh.2021.104952

Phillips, M. I., & Norgren, R. E. (1970). A rapid method for permanent implantation of an intraoral fistula in rats. *Behavior Research Methods & Instrumentation*, *2*(3), 124–124. https://doi.org/10.3758/BF03211020

Pritchard, T. C., Macaluso, D. A., & Eslinger, P. J. (1999). Taste perception in patients with insular cortex lesions. *Behavioral Neuroscience*, *113*(4), 663–671. https://doi.org/10.1037/0735-7044.113.4.663

Radovanović, M., Nanopoulos, A., & Ivanović, M. (2010). Hubs in Space: Popular Nearest Neighbors in High-Dimensional Data. *Journal of Machine Learning Research*, *11*(86), 2487–2531.

Ross, C. F., Laurence-Chasen, J. D., Li, P., Orsbon, C., & Hatsopoulos, N. G. (2024). Biomechanical and Cortical Control of Tongue Movements During Chewing and Swallowing. *Dysphagia*, *39*(1), 1–32. https://doi.org/10.1007/s00455-023-10596-9

Sadacca, B. F., Mukherjee, N., Vladusich, T., Li, J. X., Katz, D. B., & Miller, P. (2016). The Behavioral Relevance of Cortical Neural Ensemble Responses Emerges Suddenly. *Journal of Neuroscience*, *36*(3), 655–669. https://doi.org/10.1523/JNEUROSCI.2265-15.2016

Sadacca, B. F., Rothwax, J. T., & Katz, D. B. (2012). Sodium Concentration Coding Gives Way to Evaluative Coding in Cortex and Amygdala. *Journal of Neuroscience*, *32*(29), 9999–10011. https://doi.org/10.1523/JNEUROSCI.6059-11.2012

Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, *2*, e55. https://doi.org/10.7717/peerj-cs.55

Samuelsen, C. L., Gardner, M. P. H., & Fontanini, A. (2012). Effects of cue-triggered expectation on cortical processing of taste. *Neuron*, *74*(2), 410–422. https://doi.org/10.1016/j.neuron.2012.02.031

Saper, C. B. (1982). Convergence of autonomic and limbic connections in the insular cortex of the rat. *The Journal of Comparative Neurology*, *210*(2), 163–173. https://doi.org/10.1002/cne.902100207

Sasamoto, K., Nishimuta, K., Yasumatsu, K., & Ninomiya, Y. (2001). Analysis of rhythmical jaw movements produced by taste stimulation in rats. *Journal of Nutritional Science and Vitaminology*, *47*(2), 156–160. https://doi.org/10.3177/jnsv.47.156

Shammah-Lagnado, S. J., Costa, M. S. M. O., & Ricardo, J. A. (1992). Afferent connections of the parvocellular reticular formation: A horseradish peroxidase study in the rat. *Neuroscience*, *50*(2), 403–425. https://doi.org/10.1016/0306-4522(92)90433-3

Soares, E. S., Stapleton, J. R., Rodriguez, A., Fitzsimmons, N., Oliveira, L., Nicolelis, M. A. L., & Simon, S. A. (2007). Behavioral and neural responses to gustatory stimuli delivered non-contingently through intra-oral cannulas. *Physiology & Behavior*, *92*(4), 629–642. https://doi.org/10.1016/j.physbeh.2007.05.038

Spector, A. C., & Grill, H. J. (1988). Differences in the taste quality of maltose and sucrose in rats: Issues involving the generalization of conditioned taste aversions. *Chemical Senses*, *13*(1), 95–113. https://doi.org/10.1093/chemse/13.1.95

Spector, A. C., Klumpp, P. A., & Kaplan, J. M. (1998). Analytical issues in the evaluation of food deprivation and sucrose concentration effects on the microstructure of licking behavior in the rat. *Behavioral Neuroscience*, *112*(3), 678–694. https://doi.org/10.1037//0735-7044.112.3.678

Stone, B. T., Lin, J.-Y., Mahmood, A., Sanford, A. J., & Katz, D. B. (2022). LiCl-induced sickness modulates rat gustatory cortical responses. *PLoS Biology*, *20*(7), e3001537. https://doi.org/10.1371/journal.pbio.3001537

The pandas development team. (2025). *pandas-dev/pandas: Pandas* [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.16918803

Travers, J. B., & Jackson, L. M. (1992). Hypoglossal neural activity during licking and swallowing in the awake rat. *Journal of Neurophysiology*, *67*(5), 1171–1184. https://doi.org/10.1152/jn.1992.67.5.1171

Travers, J. B., & Norgren, R. (1986). Electromyographic analysis of the ingestion and rejection of sapid stimuli in the rat. *Behavioral Neuroscience*, *100*(4), 544–555. https://doi.org/10.1037//0735-7044.100.4.544

Travers, J., DiNardo, L., & Karimnamazi, H. (2000). Medullary reticular formation activity during ingestion and rejection in the awake rat. *Experimental Brain Research*, *130*(1), 78–92. https://doi.org/10.1007/s002219900223

Travers, S. P. (2002). Quinine and citric acid elicit distinctive Fos-like immunoreactivity in the rat nucleus of the solitary tract. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *282*(6), R1798–R1810. https://doi.org/10.1152/ajpregu.00590.2001

Vallat, R. (2018). Pingouin: Statistics in Python. *Journal of Open Source Software*, *3*(31), 1026. https://doi.org/10.21105/joss.01026

Venugopal, S., Travers, J. B., & Terman, D. H. (2007). A computational model for motor pattern switching between taste-induced ingestion and rejection oromotor behaviors. *Journal of Computational Neuroscience*, *22*(2), 223–238. https://doi.org/10.1007/s10827-006-0009-3

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … van Mulbregt, P. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), Article 3. https://doi.org/10.1038/s41592-019-0686-2

Zhang, C., Kang, Y., & Lundy, R. F. (2011). Terminal field specificity of forebrain efferent axons to the pontine parabrachial nucleus and medullary reticular formation. *Brain Research*, *1368*, 108–118. https://doi.org/10.1016/j.brainres.2010.10.086

Zhang, G. X., & Sasamoto, K. (1990). Projections of two separate cortical areas for rhythmical jaw movements in the rat. *Brain Research Bulletin*, *24*(2), 221–230. https://doi.org/10.1016/0361-9230(90)90209-i