Stereotyped goal-directed activity patterns in the insular cortex

Itay Talpir[1] and Yoav Livneh[1*]

[1]Department of Brain Sciences, Weizmann Institute of Science, Rehovot 76100, Israel

*Correspondence: yoav.livneh@weizmann.ac.il

- InsCtx activity manifold structure is stereotyped across different mice and motivations.
- InsCtx activity dynamics within the manifold are stereotyped across different mice and motivations.
- Stereotyped activity dynamics reflect goal-directed reward anticipation.
- Stereotyped activity dynamics do not reflect licking, taste, or valence.

Abstract

The insular cortex is involved in diverse processes including bodily homeostasis, emotions, and cognition. Yet little is known about how it processes information at the level of neuronal populations, precluding an understanding of the computations it performs. We leveraged recent advances in machine learning to study insular cortex population activity patterns (i.e., neuronal manifold) in mice performing goal-directed behaviors. We find that the insular cortex activity manifold is remarkably consistent across different animals and under different motivational states. Activity dynamics within the neuronal manifold are highly stereotyped during rewarded trials, enabling robust prediction of trial outcomes across different mice, and across various natural and artificial motivational states. Comparing goal-directed behavior with self-paced free consumption, we find that the stereotyped activity patterns reflect goal-directed reward anticipation, and not licking, taste, or valence. These findings reveal a core computation in insular cortex that could explain its involvement in pathologies involving aberrant motivations.

Introduction

The insular cortex (or insula) is involved in a variety of processes, ranging from basic functions including bodily homeostasis, emotions, motivation, decision making, social interactions, and sensory processing, to maladaptive conditions including drug addiction, chronic pain, anxiety, obesity, and eating disorders[1–8][5–8]. Yet we lack a mechanistic understanding of the core computations in InsCtx that underlie its contribution to so many diverse processes.

A central path to understanding the core computations within and across brain regions lies in measuring and making sense of neural activity of large populations of neurons in different behavioral contexts[1–3]. Indeed, there has been a recent explosion of artificial intelligence and machine learning approaches to deciphering neural activity at the level of large neuronal populations[4–6]. These approaches have begun to provide insights into the core computations of neural circuits, such as motor cortex control of limb movement, thalamic representation of head direction, and entorhinal cortex representation of spatial position[7–11]. And yet, most animal model studies of InsCtx have used either bulk signal recordings like fiber photometry, or electrical recordings from relatively small populations, not yet leveraging these new approaches to better understand InsCtx computations. These studies have indeed substantially advanced our understanding of the bulk activity of InsCtx, of the activity of individual InsCtx neurons, and also of activity patterns in small populations. Nevertheless, a comprehensive unbiased investigation of activity patterns of large InsCtx populations, taking advantage of recent developments in machine learning tools, has yet to be performed.

Previous work has suggested that InsCtx activity on short-timescales (~seconds) represents multi-modal gustatory and interoceptive sensory stimuli[12–17], as well as multi-modal salient external cues[15,18–25]. Further work has suggested that such activity may also convey emotional, motivational and valence related information[26–31]. Longer timescale changes in InsCtx activity (minutes to hours) have been suggested to represent slow changes in the physiological state of the body[20,26,32–35], as well as specific positive and negative emotional states[26–29]. However, seeing as InsCtx integrates many different inputs and is involved in many diverse processes, understanding its function at the neuronal population level requires comprehensive analyses of the repertoire of its activity

patterns (i.e., the neuronal manifold) and their dynamics. Doing so would help answer fundamental questions that remain unanswered. For example, how constrained are InsCtx population activity patterns by intrinsic factors vs. extrinsic factors such as the current behavioral context? How are different external and internal behavioral variables represented concomitantly in InsCtx population activity patterns? Are there core computations that InsCtx performs across different contexts? If so, will they have distinct signatures in population activity space?

Here, we leverage recent advances in machine learning and topological data analysis to perform an unbiased investigation of InsCtx population activity during goal-directed behaviors. We use these approaches to analyze new as well as previously published data [25,33]. We find that the repertoire of InsCtx population activity patterns (i.e., neuronal manifold) has a remarkably consistent structure across different animals, physiological needs, and behavioral contexts. The InsCtx activity manifold can be partially explained by experimentally measured external and internal behavioral variables. Moreover, by tracking the dynamics of activity within the manifold, we uncover previously unknown stereotyped patterns of activity. These stereotyped activity dynamics were highly conserved and allowed decoding across individuals, across different physiological needs, different motivational states, and different available rewards. A comparison of population dynamics during task-structured goal-directed behavior vs. self-paced, free consumption behavior revealed that the stereotyped activity patterns reflect task-dependent reward anticipation, as opposed to specific behavioral actions, taste, or valence. Our findings demonstrate the power of comprehensive population level analyses to reveal core computations of insular cortex.

Results

Insular cortex population activity manifold structure is stereotyped across different mice and motivational states

To investigate how neuronal activity patterns are structured in the insular cortex (InsCtx), we analyzed two-photon calcium imaging data from new datasets, as well as previously published ones[33]. These data were collected while thirsty (water restricted) mice performed an operant go/no go visual discrimination task in which licking following presentation of three initially arbitrary visual cues (drifting gratings) leads to rewarding (water), aversive (1 M NaCl), or neutral outcomes (Figure 1a-b)[33]. We have previously shown that most InsCtx neurons respond to some aspect of this behavioral task with either increases or decrease in activity (e.g., visual cues, licking, reward consumption)[33]. Therefore, in order to detect any underlying structure within the observed activity patterns (i.e., the "activity manifold”), we aimed to capture the temporal similarities across the entire neuronal population (Figure 1c-d).

Previous studies have shown that activity manifolds are of low dimensionality during relatively simple behavioral tasks and are thus determined by relatively few latent variables, as compared to the number of observed neurons[2,5,7,8,36–39]. However, because InsCtx is a central hub that integrates multimodal information from somatosensory, auditory, visceral, gustatory, and other limbic regions[40,41], we wondered to what extent InsCtx activity would be similarly low dimensional. Our initial goal was to determine the intrinsic dimensionality of the InsCtx population data, i.e., the minimal number of continuous variables needed to parametrize the activity manifold[36,42] (Supp. Fig. 1a-b[42]). We found that across all of our datasets, the activity exhibited a consistent intrinsic dimensionality of 6 ± 1 dimensions (mean ± SD), independent of the specific parameters used to assess dimensionality and of the number of observed neurons (Figure 1e; Supp. Fig. 1c-e). This consistently low intrinsic dimensionality, together with the diverse array of multi-modal inputs that InsCtx receives, suggest a large degree of convergence and integration of information in InsCtx during goal-directed behavior (but not in other cortical areas, see below).

Building on this finding, we used non-linear dimensionality reduction techniques to unveil the low dimensional structure of InsCtx activity patterns[4,5,7–9,38,39,43]. We selected the

target dimensionality based on the mean estimated intrinsic dimensionality across all datasets. This analysis revealed remarkably consistent structure of neuronal activity across datasets and across different mice (Figure 1d). To assess this structure similarity quantitatively, we used topological data analysis, which is emerging as an important approach for quantitative description of the structure of complex data[4,7,44–46]. We measured the similarity by assessing the difference (i.e., divergence) between the distributions of topological features across datasets (see detailed illustration of the procedure in Supp. Fig. 2a). Notably, this analysis revealed that the structure of low-dimensional activity for a dataset remained consistent across different mice and days. Structural similarity was significantly lower for shuffled datasets (Figure 1f). We further validated these findings using pairwise permutation tests[47] (Supp. Fig. 2b-c).

We next assessed whether InsCtx dimensionality and manifold structure are conserved across different motivations. To do this, we compared data from thirsty (water-restricted) mice engaged in a behavioral task to obtain water rewards, to data from hungry (food-restricted) mice engaged in the same task but to obtain liquid food rewards[25] (Figure 1g) The intrinsic dimensionality of the activity manifold in hungry mice and thirsty were quite similar (5.7 ± 1 vs. 6 ± 1; mean ± SD; Figure 1h). Moreover, the structure of the activity manifold in hungry mice was similar to that observed in thirsty mice (Figure 1i).

We next asked whether the consistent dimensionality and topological similarity we observed in InsCtx reflect behavioral constraints on brain-wide activity, i.e., whether they are unique characteristics of InsCtx. We compared the InsCtx activity manifold with the activity manifold of visual cortical areas of food-deprived mice performing the exact same behavioral task (primary visual cortex and postrhinal cortex[48]). Of note, previous work has shown that population activity patterns in visual areas, including primary visual cortex, are heavily influenced by ongoing task-related and “spontaneous” behaviors[49–51]. We found that the dimensionality and structure of the activity manifold in visual areas were both significantly higher and less consistent than those in InsCtx (Figure 1j-k). This was further confirmed by examining the distribution of P-values resulting from pairwise comparisons of the structure of activity manifolds between datasets within each group (Supp. Fig. 2d). These results suggest that the consistent dimensionality and topological similarity in

145 InsCtx are not a result of brain-wide constraints on behavior, and could be more specific
146 to a subset of brain regions that include InsCtx.

147 Collectively, our findings indicate that the neuronal activity manifold within InsCtx is low
148 dimensional and is structurally similar across mice engaged in goal-directed behaviors.
149 The consistent manifold structure across animals, suggests common activity patterns,
150 and therefore perhaps also common encoded information. We next sought to investigate
151 the dynamics of activity within this stereotyped manifold and how they relate to ongoing
152 behavior.

153

154 **Insular cortex activity dynamics within the manifold are stereotyped across**
155 **different mice**

156 Based on previous work[18,25,26,30,33], we expected InsCx activity patterns on both short and
157 long time-scales to correlate with external and internal behavioral variables (e.g., different
158 trial types, rewarding or aversive outcomes, changing satiety levels, arousal levels, etc.).
159 We color-coded the manifold based on different experimental parameters (Figure 2a, left
160 and middle; Supp. Fig. 3a) and found that some time-points that were associated with a
161 given variable were spatially clustered within the neuronal activity manifold, while others
162 were more uniformly distributed. We developed an automated method to cluster the
163 activity patterns within the manifold (termed here as "Clustered structure"). This method
164 identified different clusters of activity patterns that diverged from the center of the
165 manifold. We refer here to these clusters as ‘manifold clusters’ and refer to the center that
166 was not part of any distinct cluster as the ‘central point cloud’. (Figure 2a, right; Supp. Fig.
167 3b-c). Note that different cluster indices (represented by different colors in the figure) are
168 randomly assigned during the clustering procedure, and therefore do not necessarily map
169 to the same behavioral variable across different mice (see more below).

170 Using the clustered structure approach, we quantitatively assessed the relationship
171 between the activity manifold and external/internal behavioral variables by computing the
172 mutual information between them (Supp. Fig. 3d-e; see Methods). Three variables had
173 higher mutual information with the activity manifold than others: (1) pupil size in inter-trial
174 intervals (as a proxy for ongoing arousal levels in between epochs of overt behavior[52]),

175 (2) cumulative water rewards (as a proxy for slow changes in physiological state, i.e.,
176 water satiety[33]), and (3) trials in which the mice correctly responded to reward predicting
177 cues to receive water rewards (referred to hereafter as “rewarded trials”; Figure 2b).
178 Moreover, the mutual information for these three variables was significantly higher than
179 what was obtained for shuffled data (Figure 2c). These results are consistent with
180 previous work showing changes in InsCtx activity that are associated with rewards,
181 arousal, and physiological state[25–30,33]. We have previously shown that water rewards
182 shift population activity along a linear projection on a thirsty-quenched axis[33]. However,
183 manifold clusters associated with the quenched state were different from those
184 associated with rewarded trials (Supp. Fig. 4a-b), suggesting that there are also distinct
185 representations of water satiety and reward in InsCtx population activity.

186 We next sought to understand the dynamics associated with behavioral variables within
187 the activity manifold. We represented each activity pattern by its manifold cluster, and
188 then examined the temporal sequence of clusters during different trial types (Figure 2d).
189 InsCtx neuronal populations displayed somewhat stereotyped patterns of cluster
190 transitions during rewarded trials, characterized by movement from the central point cloud
191 to a sequence of transitions between the same 2-3 manifold clusters. To evaluate the
192 consistency of these dynamics, we calculated the pairwise similarity of cluster sequences
193 across all trial types. Cluster sequences were highly similar across different rewarded
194 trials, and substantially more so than all other trial types (Figure 2e). We confirmed these
195 results using a different dimensionality reduction method (Isomap[53]; Supp. Fig. 5a-b). We
196 also directly compared the new cluster sequence method to our previous method of
197 projecting the activity onto empirically defined linear thirsty-quenched axes in population
198 activity space[33]. As expected, we found that the linear projection naturally captures less
199 of the variance of activity within the manifold and is much less consistent across trials
200 (Supp. Fig. 4c-d).

201 We next examined to what extent trial outcomes could be decoded from the stereotyped
202 sequence of activity patterns across days and mice. The challenge in decoding across
203 mice is that our clustering technique does not yield functionally consistent cluster indices
204 across mice. That is, clusters associated with the same behavioral variable would not
205 have the same label across mice. We therefore developed a method for translating cluster

206 labels between data sets in order to properly align the activity dynamics between different
207 datasets (different days, different mice) for decoding. We first determined the probability
208 of transitions between different clusters in each individual dataset (Figure 2f, Transition
209 matrix). We then translated the cluster labels from one “reference dataset” to relabel the
210 clusters transition probabilities of the other dataset (Figure 2f, Sorted transition matrix).
211 Finally, we used the translated cluster labels to train the decoder and tested it on the other
212 “reference dataset” (Figure 2f). Using this approach we could classify rewarded trials
213 across different days, and even across different mice, with high accuracy. We achieved
214 an average accuracy of 75% ± 13% and 65% ± 13% (mean ± SD), respectively, which
215 was 4-5 times higher than chance (~12-15%). The decoding accuracy for other trial types
216 did not significantly surpass that obtained from randomly shuffling trial labels (Figure 2g).
217 We further validated these results using the P-values derived from the shuffle distributions
218 for each pairwise decoding (Supp. Fig. 6).

219 Importantly, decoding trial types on single trials across mice using this method was
220 substantially more accurate than using average activity levels across all neurons, or using
221 our previous approach of projecting activity on linear axes[33] (Supp. Fig. 6). These
222 analyses suggest that decoding accuracy for rewarded trials results from the stereotyped
223 *patterns* of activity and does not trivially reflect global changes in activity. These findings
224 demonstrate that there are highly stereotyped activity patterns within the InsCtx activity
225 manifold during rewarded trials, both across days and across mice.

226

227 **Insular cortex activity dynamics within the manifold are stereotyped across**
228 **different motivational states and rewards**

229 Given the remarkably stereotyped activity patterns we observed for rewarded trials in
230 thirsty mice (motivated by water rewards), and the consistent InsCtx manifold structure
231 during both hunger and thirst, we next tested whether similar dynamics were also evident
232 in mice motivated by food rewards. We first visualized the temporal sequence of manifold
233 clusters of rewarded trials in hungry mice in the InsCtx and found that it displayed similar
234 stereotyped cluster sequences as in thirsty mice (Figure 3a). In contrast, activity
235 sequences in visual cortical areas were very different and, as expected, primarily showed
236 activity during the presentation of the visual cue (0-2.5 seconds into the trial), independent

of reward. We then tested whether we could decode trial outcome across thirsty and hungry mice, receiving water and food rewards, respectively. The trial outcome in hungry mice could be decoded with high accuracy of 67 ± 23% (mean ± SD; vs. ~12-15% chance, see Methods) when we trained the decoder on InsCtx activity dynamics from thirsty mice. However, the decoding accuracy was similar to chance (11 ± 12%) when the decoder was trained on activity from visual areas (Figure 3b). The consistent activity dynamics in InsCtx across hunger and thirst suggests that these activity patterns do not reflect the pure sensory attributes of the reward or the underlying physiological need.

We performed further analyses to test the extent to which these activity patterns remain stereotyped in mice performing similar behavioral tasks across additional different natural and artificial motivational states. Specifically, we examined InsCtx activity from mice in which artificial hunger or thirst states were induced by chemogenetic activation of hypothalamic AgRP "hunger neurons" or SFO^{GLUT} "thirst neurons", which are the primary sensors of physiological imbalances and actuators of relevant behavioral and physiological responses[54–57]. We also analyzed InsCtx activity from mice in which we induced hyper-osmotic thirst by administration of hypertonic saline[58]. These 3 datasets gave us the unique opportunity to compare, within each mouse, the activity during the natural motivation driven by physiological need, with the artificially induced motivation.

We first visualized the sequences of manifold clusters in these mice and compared them during the natural motivational state to those after induction of the artificial motivational state. Sequence transitions during rewarded trials appeared remarkably similar in both the natural and artificially induced motivational states (Figure 3c,f,i). We quantified this using the pairwise trial similarity before and after the induction of the artificial motivation. For all motivational states, the pairwise trial similarities remained very similar between the natural and artificially induced motivational states, with no significant changes observed (Figure 3d,g,j). We therefore trained the decoder on datasets from hungry and thirsty mice that were not part of the artificial motivation experiments and then used these to test decoding in the same mice before and after artificial induction of a motivational state. Decoding accuracy for rewarded trials in artificial motivational states was similar to the natural motivational states, and both were significantly higher than chance (Figure 3e,h,k). These findings collectively demonstrate that the stereotyped activity patterns in

268 InsCtx are not specific to a particular reward (including its sensory properties), nor to the
269 specific underlying motivational state.

270

271 **Insular cortex stereotyped activity dynamics reflect goal-directed reward**
272 **anticipation, but not licking, taste, or valence**

273 The stereotyped activity patterns we observed in InsCtx across rewarded trials with
274 different rewards and motivations could reflect goal-directed behavior, or it could simply
275 reflect licking, valence, or fulfilment of a need. We next tested these different options.

276 In both hungry and thirsty mice, we observed that the onset of the stereotyped sequence
277 of activity clusters during rewarded trials appeared to be linked with anticipatory licking
278 during the trial, as evidenced by sorting trials based on the onset of anticipatory licking
279 (Figure 4a). We thus quantified the correlation between the onsets of sequential activity
280 and anticipatory licking in individual trials (Figure 4b). These correlations were notably
281 high in hungry mice (0.5 ± 0.13; mean ± SD) and in thirsty mice (0.47 ± 0.15), with all
282 correlations being statistically significant (Figure 4c). Additionally, we observed that the
283 onset of anticipatory licking consistently preceded the onset of sequential cluster activity,
284 and occurred independently of when the reward was delivered (Figure 4d-e). Together
285 with similar activity dynamics across food and water rewards (Figure 3), this further
286 supports a dissociation from the sensory aspect of the reward. Thus, these stereotyped
287 activity patterns could potentially reflect licking per se, valence (independent of the
288 reward's sensory properties), or cue-driven goal-directed behavior.

289

290 To distinguish between these possibilities, we examined trials in which mice incorrectly
291 licked to cues that do not predict reward ("false alarm" trials). We found that cluster
292 sequences in false alarm trials were more similar to rewarded trials than other trial types
293 (Figure 4f). To further investigate this, we compared cluster sequence dynamics for false
294 alarm trials and rewarded trials. We sorted false alarm trials based on the similarity of
295 *licking* pattern to rewarded trial licking (Figure 4g). Indeed, false alarm trials with similar
296 licking patterns to rewarded trials also displayed similar cluster sequence dynamics. To
297 quantify this, we assessed the correlations between licking similarity and cluster

sequence similarity in false alarms and rewarded trials within individual datasets. We observed positive correlations in 9/11 datasets (Figure 4h). Because mice do not actually receive reward in “false alarm” trials, these results further support the conclusion that InsCtx stereotyped activity patterns do not reflect reward or valence. Nevertheless, they could still potentially reflect licking behavior, or cue-associated goal-directed reward expectation.

To further distinguish between these possibilities, we compared cluster sequences when hungry mice consumed rewards while actively engaged in the operant visual discrimination task, and when they did so while freely licking in the same physical context (head-fixed, visual cue monitor is on but does not present cues) to receive rewards in the absence of predictive cues or any task structure. Notably, in both conditions mice had a similar physiological need, fulfilled using the same action (licking), to obtain the same reward. We constructed ‘pseudo-trials’ during the free reward consumption epochs by identifying significant breaks between lick bouts and analyzed the corresponding sequences of manifold clusters. As expected, average population activity increased during reward consumption in both task-structured trials and self-paced consumption (Supp. Fig. 7). However, activity dynamics during task engagement and free reward consumption were distinctly different (Figure 4i). Quantitatively, the pairwise trial similarity between operant trials and free reward consumption trials was significantly lower than the pairwise similarity within operant trials. (Figure 4j, Supp. Fig. 7). These results could be explained by differences in patterns of licking (Supp. Fig. 7). Taken together, these results suggest that the stereotyped sequential activity pattern we observed in InsCtx is closely related to learned anticipation of reward during goal-directed behavior, rather than licking, the sensory aspects of the reward, general reward, valence, or specific motivational states.

Discussion

We used unbiased topological and machine learning approaches to investigate population activity in InsCtx. We discovered that InsCtx population activity had consistent dimensionality and manifold structure across days, across different mice, and across different motivational states. Dynamics within the manifold were remarkably stereotyped, allowing us to decode trial outcomes with high accuracy across different animals, motivations, and rewards. These stereotyped dynamics reflected goal-directed reward anticipation, rather than general reward, valence, taste, or licking. We propose that such stereotyped activity patterns could reflect a core computational function of InsCtx across the many processes it is thought to be involved in.

We found that the InsCtx activity manifold (i.e., the repertoire of activity patterns), is constrained by the current behavioral context. Importantly however, we found that the same behavioral context does not yield the same manifold structure in other cortical regions involved in the behavior (e.g., visual cortex), suggesting a conjunction of two constraints that determine manifold structure – behavior and brain region. Several groups have recently reported that activity related to movement and behavior is prevalent and dominant across many brain regions[49–51,59]. Because we found different manifold structure in InsCtx and visual cortex during the same behavior, our results suggest that while behavior does constrain manifold structure, it does so in conjunction with more localized and potentially unique functions of distinct subnetworks of different brain regions.

The InsCtx activity manifold was consistent across individuals and was partially correlated with external and internal behavioral variables, which included pupil size (as a proxy for arousal behavior[52]), cumulative rewards (as a proxy for satiation[33]), and reward consumption. Our analyses were limited by the behavioral variables that were measured in our studies. It remains to be tested whether other parts of the manifold could be determined by the intrinsic connectivity of the network, as has been previously suggested[60], and/or by other variables we did not measure in these experiments[5]. These variables could include interoceptive changes, such as changes in blood pressure, gastric stretch, and blood oxygenation, all previously shown to be represented by single InsCtx

356 neurons or bulk activity[12,27]. Future studies comparing InsCtx activity across different
357 behavioral contexts could assess to what extent the intrinsic dimensionality and structure
358 of the manifold depend on specific ongoing behaviors and sensory inputs.

359

360 Most of the cellular level neural activity recordings in the InsCtx of behaving animals have
361 focused on taste processing[13–15,17]. These studies investigated relatively fast activity
362 changes on the order of hundreds of milliseconds, revealing sequential temporal coding
363 of tactile, chemosensory, and palatability information in individual neurons[13]. At the level
364 of small populations of 5-10 neurons, InsCtx activity has been shown to exhibit metastable
365 dynamics, transitioning between different states that can reflect different behavioral
366 variables, each lasting tens to hundreds of milliseconds[60–62]. Other work using activity
367 manipulations and bulk activity recordings has shown that InsCtx activity is related to
368 taste, palatability, motivation, valence, and anxiety[19,26,29–31,63,64]. Some of these variables
369 are likely represented by slower changes across the activity manifold. Nevertheless, the
370 faster stereotyped changes we discovered appear not to be related to licking, taste
371 valence, or general reward. Interestingly, a recent study found that activity of anterior
372 InsCtx projections to the brainstem encodes motivational vigor[30]. Notably, the stereotyped
373 activity patterns we described here in mid InsCtx likely do not reflect motivational vigor
374 per se, as they are sensitive to the behavioral task structure, and do not appear to change
375 with partial satiation, which reduces motivational vigor[30]. Nevertheless, seeing as InsCtx
376 activity appears to be associated with several different types of information, it will be
377 important to understand how it can simultaneously represent these different types of
378 information. Furthermore, it will be important to understand to what extent these different
379 types of information are eventually integrated before being relayed to downstream brain
380 regions to affect behavior and bodily physiology. Indeed, different InsCtx projection
381 populations (e.g., projecting to striatum, thalamus, or brainstem[40,41,65]) may convey
382 different types of information to downstream regions, and these in turn could differentially
383 weigh the input from different neurons to simultaneously extract different types of
384 information.

385

386 We previously used linear projections in activity space to suggest that reward
387 consumption transiently shifts activity towards a future satiety state[33]. Here, we used more

comprehensive analyses of the entire activity manifold and find stereotyped goal-directed reward anticipation activity patterns, which are mostly independent of changes in ongoing activity that reflect physiological state. This further underscores the importance of using both hypothesis-driven, as well as unbiased approaches to explore multi-dimensional neural activity. Thus, while our previous work provided support to a dominant model of interoception[66,67], our analyses here highlight a potentially more general computation InsCtx performs, which could be important for many behaviors InsCtx is involved in, which do not necessarily directly involve interoception.

Activity dynamics on the manifold were remarkably consistent across individuals and motivations, suggesting that these dynamics may reflect an important fundamental aspect of InsCtx function. We speculate that such goal-directed reward anticipation is related to the consistent involvement of InsCtx in cravings during pathological conditions associated with aberrant reward processing, including obesity, binge eating disorder, and drug addiction[65,68–71]. InsCtx has been suggested to be involved in these conditions, as well as in many others, by encoding the anticipated interoceptive value or outcome of actions, which when goes awry can lead to aberrant food or drug cravings[69]. We thus speculate that the stereotyped goal-directed reward anticipation patterns we discovered reflect this fundamental computation in InsCtx. As such, future work using the same analytical framework in mouse models of pathological conditions, such as obesity and drug addiction, could test this idea to form a deeper understanding of the role of InsCtx in pathological conditions that involve aberrant motivations.

Acknowledgements

We thank Mark Andermann, Alon Rubin, Yarden Cohen, Sasha Devore, Stav Shtiglitz, Ayal Lavi, Yael Prilutski, Einav Litvak, Omer Izhaki, Inbar Perets, Daniel Deitch, Tom Talpir, Omer Richmond, Gal Elyasaf, Birgit Jickeli, and Daphna Nachmani for fruitful discussions and helpful comments on the manuscript. YL is supported by research grants from the Israel Science Foundation (#860/21), the European Research Council Starting Grant (ERC StG #101039145), and the Center for New Scientists at the Weizmann Institute of Science.

Methods

Datasets analyzed for this paper were both previously published data[25,33,48], as well as unpublished data. Unpublished datasets included experiments similar to those described in the published datasets (e.g., in water or food restricted mice performing an operant visual discrimination task for water or Ensure rewards), as well as imaging during head-fixed satiation by Ensure consumption and induction of hyperosmotic thirst by intraperitoneal injection of hypertonic saline. We briefly describe the experiments below, see also refs[25,33] for more detailed descriptions.

Behavioral task

To perform the behavioral task, mice were water restricted to ~80% of their pre-restriction body weight or food-restricted to ~85% of their free-feeding body weight. We trained mice to discriminate between square-wave drifting gratings differing in orientation presented on an LCD screen (2 Hz and 0.04 cycles/degree, full-field square wave drifting gratings, 80% contrast; food cue: 0°, aversive cue: 270°, neutral cue: 135°[25,33]. All drifting gratings were presented for 2 sec, after which the mouse had a 2 sec window to respond with a lick. Licking during the visual cue was not punished, but also did not trigger delivery of the water/Ensure/salt-water/quinine. Only the first lick (if any) occurring during the response window triggered delivery of water/Ensure/salt-water/quinine. The lickspout was designed with two adjacent lick tubes (one for each outcome), such that the tongue contacted both tubes on each lick, which served as an effective deterrent for lick responses following aversive cues. Well-trained mice had a high rate of correct water/food cue licking responses (criterion: >80% of trials, usually ~90-95%), and a low rate of licking following aversive cue presentations (criterion: <50%, usually ~20-30%). Each water reward was a ~2-3 μL drop, and Ensure reward was a ~5 μL drop (0.0075 calories). Behavioral training was performed using MonkeyLogic[72] and MonkeyLogic2 (https://monkeylogic.nimh.nih.gov/).

Surgical procedures

Stereotaxic injections AAV8-CaMKII-hM3Dq-mCherry, AAV8-hSyn-DIO-hM3Dq-mCherry, and AAV1-hSyn-GCaMP6f, as well as and implantation of microprisms (2 mm

450 prisms; #MCPH-1.0; Tower Optical; coated with aluminum along their hypotenuse) were
451 performed as previously described[25,33].

452 **Two-photon imaging across different natural and artificial states**

453 Two-photon imaging of GCaMP6f was performed using a resonant-scanning two-photon
454 microscope with tiltable scanhead (Neurolabware; 31 frames/second; 1154x512 pixels).
455 All imaging was performed with a 20x 0.45 NA air objective (Olympus) with a 540 x 360
456 µm2 field of view. All imaged fields of view (FOV) were at a depth of 90-200 µm below
457 the pial surface, using a Mai Tai DeepSee laser (Newport Corp.) with laser power at 920-
458 960 nm of 35-80 mW at the front aperture of the objective (power at the sample was likely
459 substantially less due to partial transmission via the microprism). Imaging depth was
460 adjusted in between runs (every 30 min) to account for slow drift in the z plane (< 7 µm).

461 **Imaging across thirsty and quenched states**

462 We imaged mice during gradual water satiation during consecutive 30-minute runs until
463 the mice voluntarily stopped performing the task. We then performed one more imaging
464 run, the 'quenched' run, in which mice did not respond to the water cue.

465 **Imaging across hungry, satiation, and sated states**

466 We imaged mice in two blocks of trials within a session, one block during food restriction
467 and a subsequent block immediately following re-feeding. At the start of each imaging
468 session, food-restricted mice performed the visual cue discrimination task. After ~180
469 trials (30-min imaging run), we provided the mouse with ad libitum access to Ensure until
470 voluntary cessation of consumption. Ensure consumption lasted 45-75 minutes. We
471 triggered delivery of Ensure with every lick, but with a minimum inter-trial interval of 2.5 s
472 between Ensure deliveries. During this period of time, mice consumed ~3-5 mL of Ensure
473 and then voluntarily stopped licking for rewards. We then imaged additional ~180 trials
474 (30 min imaging run) while mice were satiated (operationally defined as the absence of
475 voluntary licking).

476 **Imaging during chemogenetic activation of SFO^{GLUT} neurons**

477 Following imaging during the quenched state (see above), we injected 150 µl 0.9% of
478 either saline or CNO (5 mg/kg), waited 10-15 min and started another imaging run (~180

479 trials, 30 min). For every mouse used for these experiments, we used postmortem
480 histology and immunohistochemistry to verify hM3Dq-mCherry expression in the SFO.

481 **Imaging during following injection of hypertonic saline to induce hyperosmotic**
482 **thirst**

483 Following imaging during the quenched state, we injected ~200 μL of 2M NaCl, waited 5-
484 10min and started another imaging run (~180 trials, 30 min). Mice usually re-engaged in
485 the behavioral task immediately or after 2-3 minutes.

486 **Imaging during chemogenetic activation of AgRP neurons**

487 Following imaging during hunger and satiety, mice were returned to their home-cage with
488 ad libitum access to regular chow. The next morning, we imaged the same InsCtx field of
489 view in this satiety state (100-110% of normal body weight) during ~180 trials (30 min.).
490 We then injected CNO (1-3 mg/kg). Ten minutes later, we initiated an additional imaging
491 run of ~180 trials. For every mouse used for these experiments, postmortem histology
492 and immunohistochemistry confirmed hM3Dq-mCherry expression in the hypothalamic
493 arcuate nucleus.

494 **Pupil videography during two-photon imaging**

495 We acquired data using a GigE Vision camera (Dalsa) with a 60 mm lens (Nikon
496 MicroNikkor) from a pre-selected region of interest around the eye ipsilateral to the LCD
497 monitor used to present visual cues (contralateral to the InsCtx microprism). Acquisition
498 of each frame (frame rate of 15.5 Hz) was triggered on every other frame of two-photon
499 acquisition (acquired at 31 Hz) using Scanbox software (Neurolabware). The pupil was
500 backlit with illumination originating from diffusion within the brain of the IR light used for
501 two-photon excitation during imaging. See below for details of data analysis.

502 **Image registration and time course extraction**

503 First, each acquired image was spatially down sampled by 2X. To correct for motion along
504 the imaged plane (x-y motion), each frame was registered to an average field-of-view
505 using efficient subpixel registration methods[73] Within each imaging session, each run (2-
506 8 runs/session) was registered to the first run of the day. Image stacks were de-noised
507 using principal components analyses (PCA) of every pixel across time, and by user

identification and removal of noise principal components (low eigenvalues; based on[74]). Cell masks and calcium activity time courses ('F(t)') were extracted using custom implementation of common methods[74]. To avoid use of cell masks with overlapping pixels, we only included the top 75% of pixel weights for a given mask, but users screened each prospective ROI and could edit the size of the mask, selectively removing the lowest probability pixels. We then excluded any remaining pixels identified in multiple masks. We manually verified that all cell masks had typical cell body morphology and size.

Fluorescence time courses were extracted by averaging the pixels within each region-of-interest ('ROI') mask. Fluorescence time courses for neuropil within a 25 μm annulus surrounding each ROI (but excluding adjacent ROIs and a protected ring surrounding each ROI) were also extracted ($F_{neuropil}(t)$: median value from the neuropil ring on each frame). Fluorescence timecourses were calculated as $F_{neuropil_corrected}(t) = F_{ROI}(t) - F_{neuropil}(t)$. The change in fluorescence was calculated by subtracting a running estimate of baseline fluorescence ($F_0(t)$) from $F_{neuropil_corrected}(t)$, then dividing by $F_0(t)$: $\Delta F/F(t) = (F_{neuropil_corrected}(t) - F_0(t))/ F_0(t)$. $F_0(t)$ was estimated as the 10th percentile of a 32 sec sliding window[25,33,48].

Preprocessing of neuronal data

For each collected dataset, we down-sampled $\Delta F/F$ traces by averaging every consecutive 15 frames (acquired at 31 Hz), resulting in ~0.5 seconds time-bins. Subsequently, we excluded outlier neurons with abnormally high $\Delta F/F$ values, calibrated separately for two different microscopes (2000% and 5000%). Following this, we Z-scored the activity time course for each individual neuron. Each dataset consisted of multiple 30-minute runs, acquired within the same day. For certain datasets, we applied separate Z-scoring to each run if substantial extended breaks occurred between runs or if there was a z-plane shift between the runs.

Intrinsic dimensionality estimation

To estimate the intrinsic dimensionality of the neuronal activity data, we employed the minimal neighborhood information technique as an estimator[42]. Briefly, the intrinsic dimension is estimated using the ratio formula:

$$intrinsic\ dimension = -\frac{\log\big(1 - F(\mu)\big)}{\log(\mu)}$$

Here, μ_i was determined for each data point i, where $\mu_i = \frac{r_{1,i}}{r_{2,i}}$, with $r_{1,i}$ and $r_{2,i}$ representing the Euclidean distances of the nearest and second-nearest data points to point i, respectively. $F(\mu_i)$ indicates the percentile of μ_i with respect to all other data points. Subsequently, μ_i was computed for each data point, and the highest 4% of points were excluded. The intrinsic dimension was defined as the slope calculated from a linear regression between $\log\big(1 - F(\mu)\big)$ and $\log(\mu)$ across all data points.

Non-linear dimensionality reduction

To reduce the dimensionality of the neuronal data matrix X, we applied Laplacian eigenmaps[75] implemented by a pre-established pipeline[76]. Laplacian eigenmaps, a spectral non-linear dimensionality reduction technique, operates on a data matrix X with dimensions $N \times T$, by generating a weighted adjacency graph W for the T high-dimensional data points such as each time point $t \in \mathbb{R}^N$. This adjacency graph is computed for all pairs of data points, where the weight between a pair of data points $x_i, x_j \in X$ is assigned as $W_{ij} = 1$ if the data points are connected and $W_{ij} = 0$ otherwise. Data points are considered connected if they fall within the K nearest neighbors of each other, where K is a user-defined parameter (see below). Following the computation of W, a diagonal weight matrix D is constructed with $D_{ii} = \sum_j W_{ji}$, which is equivalent to aggregating the rows of W. Notably, the Laplacian of the original data matrix is $L = D - W$. The process of obtaining the low-dimensional embedding is achieved by calculating the eigenvalues and eigenvectors of $L\boldsymbol{f} = \lambda D\boldsymbol{f}$. After excluding the leading eigenvector[75], the following m eigenvectors are utilized to obtain a new low-dimensional mapping for each data point x_i, which is given by[17]

$$x_i \rightarrow (f_1(i), \dots, f_m(i))$$

with m representing the desired output dimensionality. For every dataset X, we conducted two rounds[38] of dimensionality reduction using the algorithm described above. In the initial iteration, a 20-dimensional matrix was generated, with a parameter K set to 7.5% of the time points in matrix X. The second iteration was performed on the 20-dimensional matrix

obtained from the preceding step. This culminated in the matrix reaching its ultimate dimension – the mean estimated intrinsic dimension of 6 – with a parameter K set to 2.5% of the time points in matrix X. The mean estimated intrinsic dimension of 6 was derived from the process described above.

Topological features analysis and topological similarity

To assess the topological features of the data, we quantified for each dataset its components (β0) and holes (β1) as a function of the radius threshold, using a pre-established algorithm[45] (TDAStats: R pipeline for computing persistent homology in topological data analysis, https://github.com/rrrlw/TDAstats). Due to computational considerations, we determined the topological characteristics of clustered timepoints within the reduced data matrix, acquired through K-means clustering with a chosen K value of 80. Given our utilization of a stochastic variant of K-means, this process was repeated 20 times for each dataset. For every topological feature extracted from the clustered data matrix, we quantified its lifespan and derived the lifespan distribution for all topological features. The topological resemblance between a pair of datasets was defined as the mean Wasserstein distance[77] between the lifespan distribution of features from both datasets, covering all 20 iterations of K-means clustering. To ascertain the significance of differences between dataset pairs, permutation tests were conducted, following a methodology outlined in prior work[47]. The lifespan distributions of a pair of datasets were pooled and then systematically rearranged to generate two separate subsamples, adhering to the original sizes of each distribution. The Wasserstein distance was computed between these subsamples. This procedure was repeated n=200 times, and a P-value was established as the percentage of permutations yielding a Wasserstein distance less than the original distance between the dataset pairs. The ultimate P-value for every pair of datasets was determined as the average of the P-values derived from this process across all 20 K-means clustering iterations.

Clustering the activity manifold

To evaluate the activity manifold, our objective was to create a standardized comparison across datasets through an automated clustering approach. To achieve this, we devised an unsupervised clustering strategy that recognizes data points that deviate from the central point cloud, and thus classifies them as distinct neuronal clusters. To do so we

systematically excluded time points that were closest to the centroid of the reduced data matrix. Starting from 40 times points and progressing to the entire dimensionality of the data matrix. For each set of excluded timepoints, we conducted 20 iterations of K-means clustering with K values spanning from 2 to 20. The mean squared error (MSE) was computed between the resulting clusters and their centroids across all iterations and for each K value. This procedure produced a two-dimensional grid with its two dimensions being the fraction of excluded timepoints and the chosen K values for K-means clustering. The cells of the grid contained MSE values over the 20 clustering repetitions. To identify the optimal configuration, we scaled both grid dimensions to 1 and determined the grid cell closest to the origin of a three-dimensional Cartesian system formed by the grid, where the third dimension was the MSE value within each cell. This process yielded, for each dataset an optimal configuration consisting of 2 parameters: a K value, and the number of datapoints to be considered as the central point cloud. Across virtually all utilized datasets the identified optimal K value was 8, except one dataset in which it was 7. After obtaining the optimal configuration, our aim was to derive deterministic clustering labels for each dataset. This was accomplished by implementing the optimal configuration, isolating timepoints deemed part of the central point cloud, and executing 500 iterations of K-means clustering with the optimal K value for the remaining timepoints. The resulting clustering labels were aggregated into a $500 \times T$ matrix. This matrix was then subjected to an additional clustering step using the ward agglomerative complete linkage hierarchical clustering method[78]. This yielded the final labels of the clustered activity manifold.

Activity manifold parameterization

To evaluate the contribution of behavioral variables to the activity manifold, we devised labeling for the distinct behavioral variables, considering the total timepoints under analysis. These variables were categorized into two groups: discrete (binary) and continuous. Discrete binary variables encompassed the potential classes of trial outcomes within the behavioral task (e.g., response vs. no response for a given trial type). For each of the six trial outcome classes, a corresponding label vector was generated, spanning T timepoints. A value of $Trial\ outcome\ Label = 1$ was assigned for timepoints ranging from 3 to 6 seconds after the cue onset for each class, while a value of

627 $Trial\ outcome\ Label = 0$ was allocated to the remaining timepoints. This enabled the
628 isolation of timepoints exclusively linked to that specific trial outcome, excluding those
629 related to visual cues. As for continuous variables, namely Pupil Size and Cumulative
630 Consumed Rewards, their magnitude across the entire experiment was quantified. Pupil
631 size was quantified using code from previously published work[79]. Two label vectors, each
632 with a length of T, were established for these variables. To ensure equitable comparison
633 with discrete variables, both Pupil Size and Cumulative Consumed Rewards were divided
634 into eight bins.

635 **Mutual information between behavioral variables and the activity manifold**

636 To determine the mutual information between behavioral variables and the activity
637 manifold, we constructed a joint contingency table that related the two label vectors:
638 $Cluster$ and $Behavior$. Here, $Cluster$ represents a vector of size T containing different
639 neuronal cluster labels (including the central point cloud), while Behavior was a T-sized
640 vector representing behavioral variables (ranging from 0 to 1 for discrete variables and 1
641 to 8 for continuous variables). Following this, we utilized the contingency table to
642 formulate a joint probability space, denoted as $P_{Cluster \times Behavior}$. The calculation of mutual
643 information was carried out according to the formula:

644 $$MI = \sum_{c} \sum_{b} P_{Cluster \times Behavior}(c, b) * \log\left(\frac{P_{Cluster \times Behavior}(c, b)}{P_{Cluster}(c) * P_{Behavior}(b)}\right)$$

645 Where $P_{Cluster}$ and $P_{Behavioor}$ are the respective marginal probability distributions. To
646 reduce potential variations caused by different behavioral factors and ensure an equitable
647 comparison with discrete variables, for the continuous variables we exclusively included
648 time points unaffected by behavior, following established procedures from our previous
649 work[33]. These time points corresponded to the final 3 seconds of inter-trial intervals
650 following the correct rejection of aversive and neutral trials (referred to as non-responded
651 trials).

652 **Trial structure manifold dynamics**

653 To evaluate manifold dynamics in relation to the trial structure, we compiled the sequential
654 neuronal cluster label (including the central point cloud) time course for every trial. This
655 was achieved by capturing a 10-second interval from the onset of the cue for each trial

class. They yielded a sequence of 20 neuronal cluster labels, subsequently combined into an $M \times 20$ matrix. Here, M denotes the trial count, and each row signifies the sequence of visited clusters during the respective trial.

Trial similarity

To assess trial dynamics similarity, our focus was on highlighting the sequence of neuronal clusters. To achieve this, we designated the label -1 to each timepoint associated with the central point cloud. Meanwhile, other timepoints were allocated labels ranging from 1 to 7, corresponding to the respective neuronal cluster assignments. Subsequently, we created a trial structure matrix as previously outlined and computed pairwise trial similarity using cosine distance.

$$Trial\ Similarity(Trial_i, Trial_j) = \frac{Trial_i \cdot Trial_j}{\|Trial_i\|\ \|Trial_j\|}$$

Here, $Trial_i$ represents the timepoint sequence of labels in the ith trial. This procedure resulted in a $M \times M$ for each dataset. Following that, we concentrated on the lower triangle of the matrix and pooled all values across all datasets.

Decoding trial outcome across datasets

To enable across-datasets decoding, we generated a trial dynamics matrix using the same labeling approach employed for assessing trial similarity. Our goal was to enable decoding between datasets, which necessitated an alignment process. For each dataset, we initiated alignment by constructing a transition probability matrix between all neuronal clusters and the central point cloud. This process yielded a 8×8 probability matrix for each dataset, except for one dataset with 7 clusters that was excluded from this analysis. Following this, the probabilities underwent clustering through the ward agglomerative complete linkage hierarchical clustering method[78]. Using the hierarchical clustering structure, we derived a cluster order based on the similarity of their probabilities. This order served as a mask across datasets to allow translation of clustering labels among dataset pairs. Essentially, this step aimed to address the question: Given transition probabilities, which cluster in dataset A most closely aligns with a cluster in dataset B. With the labels clusters of dataset B aligned to dataset A, we proceeded to compute the pairwise trial similarity between each trial pair in both datasets. For each trial a in dataset

A, the decoded class of trial outcome was determined by maximizing the sum of Trial Similarity values over trials b in dataset B:

$$argmax_{class} \sum_{b \in B} Trial\ Similarity(Trial_{b(class)}, Trial_a)$$

This process is analogous to identifying, on average after translation, which class of trials in dataset B most closely resembles an individual trial in dataset A. To further validate the decoding accuracy, we established a surrogate shuffle distribution by randomly shuffling trial classes in both datasets 200 times. We then calculated the decoding accuracy for each shuffle and computed the P-value by determining the percentile of shuffles that exhibited a higher decoding accuracy compared to the measured decoding accuracy. This entire process was repeated for every pairwise combination of datasets.

Decoding trial outcome across datasets based on overall activity levels and thirsty-quenched linear axis

To decode trial outcomes across dataset based on overall activity levels, we calculated the overall activity levels within each trial by averaging the collective population activity across neurons. To decode trial outcomes across dataset based on the thirsty-quenched linear axis, we defined this axis for each mouse by $\bar{x}_{Thirsty} - \bar{x}_{Quenched}$, where $\bar{x}$ is the population vector of mean ongoing activity in each state. We projected peri-cue activity onto this axis by calculating the dot product of this vector with the time-varying pattern of InsCtx population activity, x(t). We then scaled values along this axis per mouse, ascribing a value of 1 when x(t) = $\bar{x}_{Thirsty}$, a value of 0 when x(t) = $\bar{x}_{Quenched}$ as 0, and intermediate values for patterns that fall between $\bar{x}_{Thirsty}$ and $\bar{x}_{Quenched}$ projected this data onto a thirsty-quenched linear axis, as outlined in reference[33]. Second, we utilized a similar decoding as explained above. Instead of training it with aligned trial labels for pairs of trials, we evaluated used the overall activity levels within each trial or the projection of the trial onto the thirsty-quenched axis as a training dataset.

Quantification of anticipatory licking onset

The onset of anticipatory licking during a trial was defined as the time point when the first licking bout took place within the response window after cue presentation. Licking bouts

were considered as instances where mice notably increased their licking rate. To detect these bouts, for each trial, we calculated the time intervals between consecutive licks occurring after cue presentation. We then identified the pair of licks with intervals shorter than a specific threshold:

$$\frac{1}{\mu_{Cue\ licks} + C * \sigma_{Cue\ licks}}$$

Where $\mu_{Cue\ licks}$, $\sigma_{Cue\ licks}$ were determined by calculating the average and standard deviation of licks based on the first 15 frames (0.5 sec) of cue presentation across all trials. The constant C varied; it was set as 0.1 for datasets involving water rewards and 0.35 for datasets involving food rewards. The time point at which the first pair of licks exceeded this threshold was considered the onset of anticipatory licking.

Quantification of cluster onset

To identify the onset of the cluster sequence, we first determined the two neuronal clusters with the highest occupancy during the response window for each dataset. Following this, we identified the initial time point within each trial where a transition occurred from the central point cloud to each of the identified clusters. The onset was defined as the first time point in each trial where the transition occurred with the cluster that exhibited the highest onset correlations with the lick onsets across all trials.

Lick similarity between pairs of trials

To define the lick similarity between pairs of trials, for each trial, we smoothed the binary licks vector $Licks_i$. To do so we applied a sliding window average of 100 frames (~3.3 sec). The smoothed lick pattern for the ith trial was calculated as follows:

$$SL_{i,a} = \frac{1}{100} \sum_{a}^{a+100} Licks_{i,a}$$

Where a is the ath licking frame in the original binary vector and SL_i is the overall smoothed licking pattern. Lick similarity between the ith and jth trials was determined using the formula:

$$Lick\ Similarity(Licks_i, Licks_j) \frac{1}{\left\| SL_i - SL_j \right\|^2}$$

740 **Construction pseudo-trials in freely consuming mice.**

741 To construct pseudo-trials of consumption in freely consuming mice, we identified time
742 points in which mice initiated a licking bout. To do so we identified all the time points within
743 the free consumption epoch of the experiment in which the interval between two
744 consecutive licks was above 1 second, unless this did not yield enough pseudo-trials, in
745 which case we used 0.5 sec as the threshold (2/5 datasets). Every pair of licks that passed
746 these criteria could be used to define the onset of a pseudo-trial. We then took all the time
747 points that occurred two seconds before and four seconds after the second lick in that
748 interval. We used these time points to identify the neuronal cluster dynamics and licking
749 pattern that took place in the experiment and constructed a pseudo-trial matrix. We then
750 removed pseudo-trials that overlapped with each other to avoid inflation of the matrix. To
751 control for behavioral differences between operant task licking and free consumption
752 licking, we repeated the analyses in (Fig. 4j-i) while selecting only pseudo-trials of free
753 consumption in which the average licking rate in the first two seconds of the lick bout were
754 within the mean ± standard deviation of the average licking rate in the first two seconds
755 of operant task licking (Supp. Fig. 7c-e).

756

757

References

1. Vyas, S., Golub, M. D., Sussillo, D. & Shenoy, K. V. Computation Through Neural Population Dynamics. *Annu Rev Neurosci* 249–275 (2020) doi:10.1146/annurev-neuro-092619.

2. Jazayeri, M. & Afraz, A. Navigating the Neural Space in Search of the Neural Code. *Neuron* vol. 93 1003–1014 Preprint at https://doi.org/10.1016/j.neuron.2017.02.019 (2017).

3. Saxena, S. & Cunningham, J. P. Towards the neural population doctrine. *Current Opinion in Neurobiology* vol. 55 103–111 Preprint at https://doi.org/10.1016/j.conb.2019.02.002 (2019).

4. Schneider, S., Lee, J. H. & Mathis, M. W. Learnable latent embeddings for joint behavioural and neural analysis. *Nature* **617**, 360–368 (2023).

5. Mitchell-Heggs, R., Prado, S., Gava, G. P., Go, M. A. & Schultz, S. R. Neural manifold analysis of brain circuit dynamics in health and disease. *Journal of Computational Neuroscience* vol. 51 1–21 Preprint at https://doi.org/10.1007/s10827-022-00839-3 (2023).

6. Langdon, C., Genkin, M. & Engel, T. A. A unifying perspective on neural manifolds and circuits for cognition. *Nat Rev Neurosci* **24**, 363–377 (2023).

7. Gardner, R. J. *et al.* Toroidal topology of population activity in grid cells. *Nature* **602**, 123–128 (2022).

8. Chaudhuri, R., Gerçek, B., Pandey, B., Peyrache, A. & Fiete, I. The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nat Neurosci* **22**, 1512–1520 (2019).

9. Ajabi, Z., Keinath, A. T., Wei, X.-X. & Brandon, M. P. Population dynamics of head-direction neurons during drift and reorientation. *Nature* (2023) doi:10.1038/s41586-023-05813-2.

10. Gallego, J. A., Perich, M. G., Chowdhury, R. H., Solla, S. A. & Miller, L. E. Long-term stability of cortical population dynamics underlying consistent behavior. *Nat Neurosci* **23**, 260–270 (2020).

11. Churchland, M. M. *et al.* Neural population dynamics during reaching. *Nature* **487**, 51–56 (2012).

12. Cechetto, D. F. & Saper, C. B. *Evidence for a Viscerotopic Sensory Representation in the Cortex and Thalamus in the Rat*. *THE JOURNAL OF COMPARATIVE NEUROLOGY* vol. 262 (1987).

13. Katz, D. B., Simon, S. A. & Nicolelis, M. A. L. *Dynamic and Multimodal Responses of Gustatory Cortical Neurons in Awake Rats*. (2001).

14. Jones, L. M., Fontanini, A. & Katz, D. B. Gustatory processing: a dynamic systems approach. *Curr Opin Neurobiol* **16**, 420–428 (2006).

15. Kadohisa, M., Rolls, E. T. & Verhagen, J. V. Neuronal Representations of Stimuli in the Mouth: The Primate Insular Taste Cortex, Orbitofrontal Cortex and Amygdala. *Chem Senses* **30**, 401–419 (2005).

16. Small, D. M. Flavor is in the brain. *Physiol Behav* **107**, 540–552 (2012).

17. De Araujo, I. E. & Simon, S. A. The gustatory cortex and multisensory integration. *International Journal of Obesity 2009 33:2* **33**, S34–S43 (2009).

18. Gardner, M. P. H., Fontanini, A., Gardner, M. P. H. & Fontanini, A. Encoding and Tracking of Outcome-Specific Expectancy in the Gustatory Cortex of Alert Rats. *Journal of Neuroscience* **34**, 13000–13017 (2014).

19. Samuelsen, C. L., Gardner, M. P. H. & Fontanini, A. Effects of Cue-Triggered Expectation on Cortical Processing of Taste. *Neuron* **74**, 410–422 (2012).

20. de Araujo, I. E. *et al.* Neural Ensemble Coding of Satiety States. *Neuron* **51**, 483–494 (2006).

21. Rodgers, K. M., Benison, A. M., Klein, A. & Barth, D. S. Auditory, Somatosensory, and Multisensory Insular Cortex in the Rat. *Cerebral Cortex* **18**, 2941–2951 (2008).

22. Stapleton, J. R., Lavine, M. L., Wolpert, R. L., Nicolelis, M. A. L. & Simon, S. A. Rapid Taste Responses in the Gustatory Cortex during Licking. *Journal of Neuroscience* **26**, 4126–4138 (2006).

23. Vincis, R. & Fontanini, A. Associative learning changes cross-modal representations in the gustatory cortex. *Elife* **5**, (2016).

24. Yamamoto, T., Matsuo, R., Kiyomitsu, Y. & Kitamura, R. Sensory inputs from the oral region to the cerebral cortex in behaving rats: an analysis of unit responses in cortical somatosensory and taste areas during ingestive behavior. *https://doi.org/10.1152/jn.1988.60.4.1303* **60**, 1303–1321 (1988).

25. Livneh, Y. *et al.* Homeostatic circuits selectively gate food cue responses in insular cortex. *Nature* **546**, 611–616 (2017).

26. Gehrlach, D. A. *et al.* Aversive state processing in the posterior insular cortex. *Nat Neurosci* **22**, 1424–1437 (2019).

27. Klein, A. S., Dolensek, N., Weiand, C. & Gogolla, N. Fear balance is maintained by bodily feedback to the insular cortex in mice. *Science (1979)* **374**, 1010–1015 (2021).

28. Dolensek, N., Gehrlach, D. A., Klein, A. S. & Gogolla, N. Facial expressions of emotion states and their neuronal correlates in mice. *Science (1979)* **368**, (2020).

29. Nicolas, C. *et al.* Linking emotional valence and anxiety in a mouse insula-amygdala circuit. *Nat Commun* **14**, 5073 (2023).

30. Deng, H. *et al.* A genetically defined insula-brainstem circuit selectively controls motivational vigor. *Cell* **184**, 6344-6360.e18 (2021).

31. Wang, L. *et al.* The coding of valence and identity in the mammalian taste system. *Nature* **558**, 127–131 (2018).

32. Egan, G. *et al.* Neural correlates of the emergence of consciousness of thirst. *Proc Natl Acad Sci U S A* **100**, 15241–15246 (2003).

33. Livneh, Y. *et al.* Estimation of Current and Future Physiological States in Insular Cortex. *Neuron* **105**, 1094-1111.e10 (2020).

34. Meier, L. *et al.* Thirst-dependent activity of the insular cortex reflects its emotion-related subdivision: a cerebral blood flow study. *Elsevier*.

35. Tataranni, P. A. *et al.* Neuroanatomical correlates of hunger and satiation in humans using positron emission tomography. *Proc Natl Acad Sci U S A* **96**, 4569–4574 (1999).

36. Jazayeri, M. & Ostojic, S. Interpreting neural computations by examining intrinsic and embedding dimensionality of neural activity. *Current Opinion in Neurobiology* vol. 70 113–120 Preprint at https://doi.org/10.1016/j.conb.2021.08.002 (2021).

37. Nieh, E. H. *et al.* Geometry of abstract learned knowledge in the hippocampus. *Nature* **595**, 80–84 (2021).

38. Rubin, A. *et al.* Revealing neural correlates of behavior without behavioral measurements. *Nat Commun* **10**, (2019).

39. Humphries, M. D. Strong and weak principles of neural dimension reduction. *Neuron Behav Data Anal Theory* **5**, 1–28 (2021).

40. Gehrlach, D. A. *et al.* A whole-brain connectivity map of mouse insular cortex. *Elife* **9**, 1–78 (2020).

41. Saper, C. B. The central autonomic nervous system: Conscious visceral perception and autonomic pattern generation. *Annual Review of Neuroscience* vol. 25 433–469 Preprint at https://doi.org/10.1146/annurev.neuro.25.032502.111311 (2002).

42. Facco, E., D'Errico, M., Rodriguez, A. & Laio, A. Estimating the intrinsic dimension of datasets by a minimal neighborhood information. *Sci Rep* **7**, (2017).

43. Deitch, D., Rubin, A. & Ziv, Y. Representational drift in the mouse visual cortex. *Current Biology* **31**, 4327-4339.e6 (2021).

44. Rybakken, E., Baas, N. & Dunn, B. Decoding of neural data using cohomological feature extraction. *Neural Computation* vol. 31 68–93 Preprint at https://doi.org/10.1162/neco_a_01150 (2019).

45. R. Wadhwa, R., F.K. Williamson, D., Dhawan, A. & G. Scott, J. TDAstats: R pipeline for computing persistent homology in topological data analysis. *J Open Source Softw* **3**, 860 (2018).

46. Rouse, T. C., Ni ID, A. M., Huang, C. & Cohen, M. R. Topological insights into the neural basis of flexible behavior. *Proceedings of the National Academy of Sciences* **120**, (2023).

47. Robinson, A. & Turner, K. Hypothesis testing for topological data analysis. *J Appl Comput Topol* **1**, 241–261 (2017).

48. Burgess, C. R. *et al.* Hunger-Dependent Enhancement of Food Cue Responses in Mouse Postrhinal Cortex and Lateral Amygdala. *Neuron* **91**, 1154–1169 (2016).

49. Salkoff, D., Zagha, E., … E. M.-C. & 2020, undefined. Movement and performance explain widespread cortical activity in a visual detection task. *academic.oup.comDB Salkoff, E Zagha, E McCarthy, DA McCormickCerebral Cortex, 2020•academic.oup.com*.

50. Musall, S., Kaufman, M., Juavinett, A., Gluf, S. & Churchland, A. Single-trial neural dynamics are dominated by richly varied movements. *Nat Neurosci* **22**, 1677–1686 (2019).

866 51. Stringer, C. *et al.* Spontaneous behaviors drive multidimensional, brainwide activity. *Science*
867 *(1979)* **364**, (2019).

868 52. McGinley, M. J. *et al.* Waking State: Rapid Variations Modulate Neural and Behavioral Responses.
869 *Neuron* **87**, 1143–1161 (2015).

870 53. Tenenbaum, J. B., de Silva, V. & Langford, J. C. A Global Geometric Framework for Nonlinear
871 Dimensionality Reduction. *Science (1979)* **290**, 2319–2322 (2000).

872 54. Augustine, V., Gokce, S. K. & Oka, Y. Peripheral and Central Nutrient Sensing Underlying Appetite
873 Regulation. *Trends Neurosci* **41**, 526–539 (2018).

874 55. Andermann, M. L. & Lowell, B. B. Toward a Wiring Diagram Understanding of Appetite Control.
875 *Neuron* **95**, 757–778 (2017).

876 56. Zimmerman, C. A. & Knight, Z. A. Layers of signals that regulate appetite. *Curr Opin Neurobiol* **64**,
877 79–88 (2020).

878 57. Sternson, S. M. & Eiselt, A. K. Three Pillars for the Neural Control of Appetite.
879 *https://doi.org/10.1146/annurev-physiol-021115-104948* **79**, 401–423 (2017).

880 58. Zimmerman, C. A. *et al.* Thirst neurons anticipate the homeostatic consequences of eating and
881 drinking. *Nature* **537**, 680–684 (2016).

882 59. Allen, W. E. *et al.* Thirst regulates motivated behavior through modulation of brainwide neural
883 population dynamics. *Science (1979)* **364**, (2019).

884 60. Mazzucato, L., La Camera, G. & Fontanini, A. Expectation-induced modulation of metastable
885 activity underlies faster coding of sensory stimuli. *Nat Neurosci* **22**, 787–796 (2019).

886 61. Jones, L., Fontanini, A., Sadacca, B., Miller, P. & Katz, D. Natural stimuli evoke dynamic sequences
887 of statesin sensory cortical ensembles. *Proc Natl Acad Sci U S A* **104**, 18772–18777 (2007).

888 62. Lang, L., Camera, G. La & Fontanini, A. Temporal progression along discrete coding states during
889 decision-making in the mouse gustatory cortex. *PLoS Comput Biol* **19**, (2023).

890 63. Mukherjee, N., Wachutka, J. & Katz, D. B. Impact of precisely-timed inhibition of gustatory cortex
891 on taste behavior depends on single-trial ensemble dynamics. *Elife* **8**, (2019).

892 64. Yiannakas, A. *et al.* Parvalbumin interneuron inhibition onto anterior insula neurons projecting to
893 the basolateral amygdala drives aversive taste memory retrieval. *Curr Biol* **31**, 2770-2784.e6
894 (2021).

895 65. Livneh, Y. & Andermann, M. L. Cellular activity in insular cortex across seconds to hours:
896 Sensations and predictions of bodily states. *Neuron* vol. 109 3576–3593 Preprint at
897 https://doi.org/10.1016/j.neuron.2021.08.036 (2021).

898 66. Barrett, L. F. & Simmons, W. K. Interoceptive predictions in the brain. *Nature Reviews*
899 *Neuroscience* vol. 16 419–429 Preprint at https://doi.org/10.1038/nrn3950 (2015).

900 67. Owens, A. P., Allen, M., Ondobaka, S. & Friston, K. J. Interoceptive inference: From computational
901 neuroscience to clinic. *Neurosci Biobehav Rev* **90**, 174–183 (2018).

68. Uddin, L. Q. Salience processing and insular cortical function and dysfunction. *Nature Reviews Neuroscience* vol. 16 55–61 Preprint at https://doi.org/10.1038/nrn3857 (2015).

69. Naqvi, N. H., Gaznick, N., Tranel, D. & Bechara, A. The insula: A critical neural substrate for craving and drug seeking under conflict and risk. *Ann N Y Acad Sci* **1316**, 53–70 (2014).

70. Craig, A. D. Interoception: the sense of the physiological condition of the body. *Curr Opin Neurobiol* **13**, 500–505 (2003).

71. Gogolla, N. The insular cortex. *Current Biology* 573–591 (2017).

72. Asaad, W. F. & Eskandar, E. N. A flexible software tool for temporally-precise behavioral control in Matlab. *J Neurosci Methods* **174**, 245–258 (2008).

73. Bonin, V., Histed, M. H., Yurgenson, S. & Clay Reid, R. Local diversity and fine-scale organization of receptive fields in mouse visual cortex. *Journal of Neuroscience* **31**, 18506–18521 (2011).

74. Mukamel, E. A., Nimmerjahn, A. & Schnitzer, M. J. Automated Analysis of Cellular Signals from Large-Scale Calcium Imaging Data. *Neuron* **63**, 747–760 (2009).

75. Belkin, M. & Niyogi, P. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Comput* **15**, 1373–1396 (2003).

76. Kraemer, G., Reichstein, M. & Mahecha, M. D. dimRed and coRanking-Unifying Dimensionality Reduction in R. *R J* **10**, 342–358 (2018).

77. Ramdas, A., Garcia, N. & Cuturi, M. On Wasserstein Two Sample Testing and Related Families of Nonparametric Tests. *ArXiv* (2015).

78. Murtagh, F. & Legendre, P. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *J Classif* **31**, 274–295 (2014).

79. Winkelmeier, L. *et al.* Striatal hub of dynamic and stabilized prediction coding in forebrain networks for olfactory reinforcement learning. *Nat Commun* **13**, (2022).

Figure 1

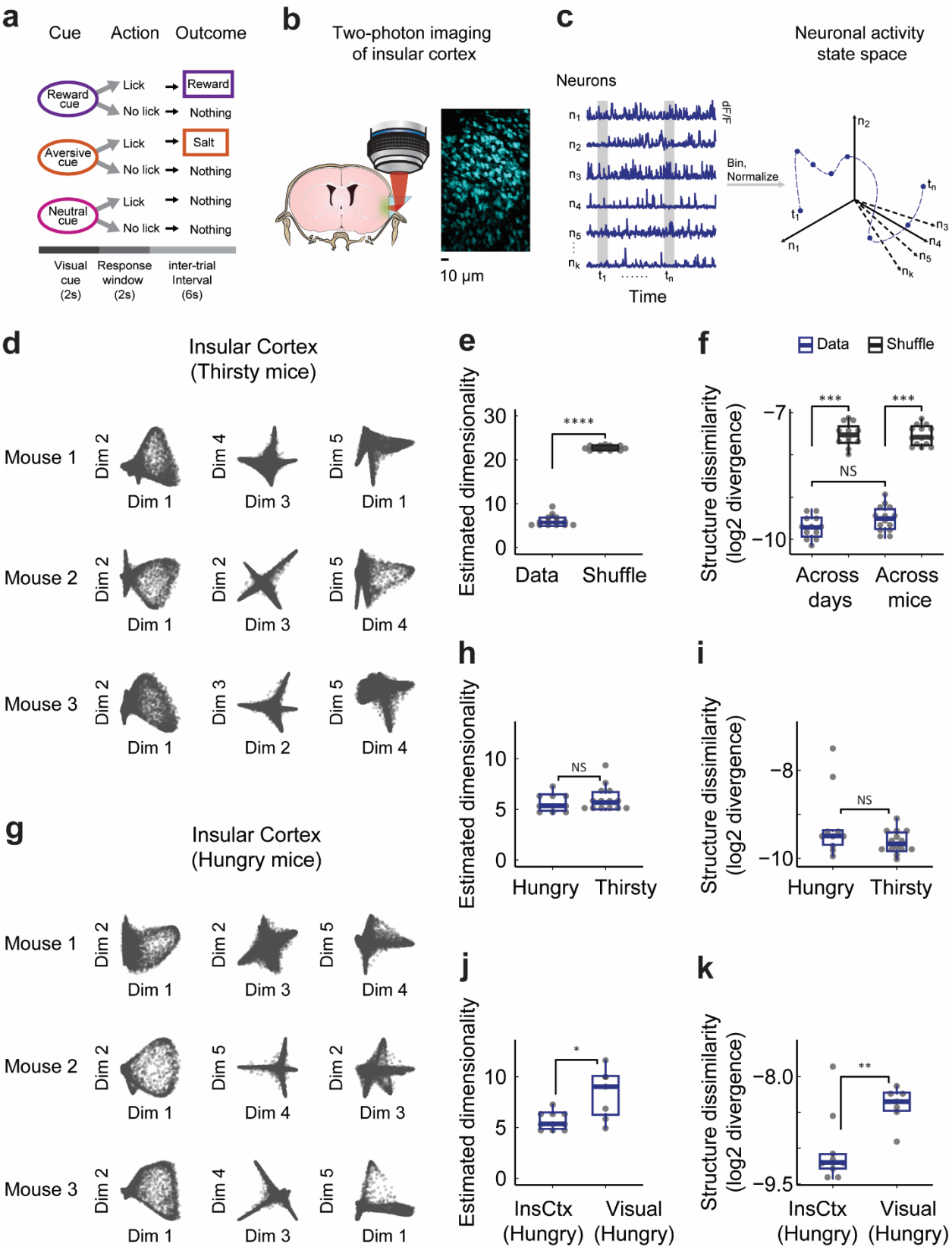

a
Cue
Action
Outcome
Reward cue
Lick
Reward
No lick
Nothing
Aversive cue
Lick
Salt
No lick
Nothing
Neutral cue
Lick
Nothing
No lick
Nothing
Visual cue (2s)
Response window (2s)
inter-trial Interval (6s)
b
Two-photon imaging of insular cortex
10 µm
c
Neuronal activity state space
Neurons
dF/F
Bin, Normalize
Time
d
Insular Cortex (Thirsty mice)
Mouse 1
Mouse 2
Mouse 3
Dim 1
Dim 2
Dim 3
Dim 4
Dim 5
e
Estimated dimensionality
****
Data
Shuffle
f
Data
Shuffle
Structure dissimilarity (log2 divergence)
***
NS
Across days
Across mice
g
Insular Cortex (Hungry mice)
h
Estimated dimensionality
NS
Hungry
Thirsty
i
Structure dissimilarity (log2 divergence)
NS
Hungry
Thirsty
j
Estimated dimensionality
*
InsCtx (Hungry)
Visual (Hungry)
k
Structure dissimilarity (log2 divergence)
**
InsCtx (Hungry)
Visual (Hungry)

Figure 1 – InsCtx activity manifold structure is stereotyped across different mice and motivational states

a. Schematic of the operant visual discrimination task.

b. Two-photon calcium imaging of InsCtx through a microprism, schematic coronal brain section and approach.

c. Illustration of the approach to study population activity in the InsCtx.

d. Three representative planes of dimensionality-reduced neuronal activity over time, from three different thirsty mice. Each point represents the population activity pattern in a 0.5 sec time bin. Note the high qualitative similarity between different mice.

e. Estimation of intrinsic dimensionality for each individual dataset vs. shuffled data. The estimated dimensionality was significantly lower than chance ($P < 7 * 10^{-5}$, one-tailed Wilcox signed-rank test).

f. Quantification of topological dissimilarity ($\beta 0$ features) between all datasets (see also Supp. Figure 2). Notably, there was a significant difference observed when comparing either across days or across mice to chance levels ($P < 0.0003$, one-tailed Wilcox signed-rank test), while comparisons of the topological dissimilarity across days and across mice yielded no significant differences ($P > 0.1$, two-tailed Wilcox rank sum test).

g. Same as in 'd'., for data obtained from hungry mice. Different mice than those in 'd'.

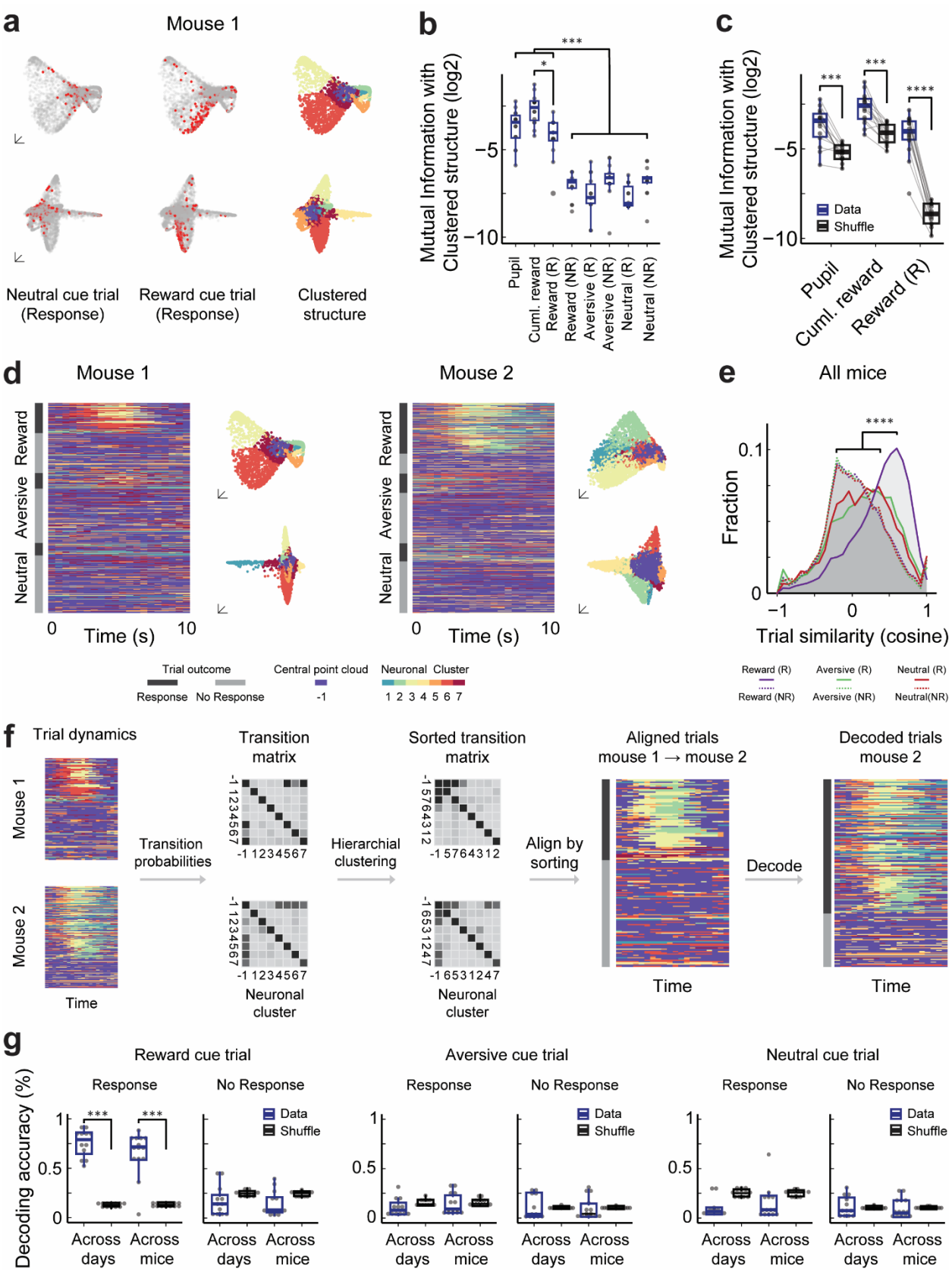
h. Similar estimated intrinsic dimensionality across hungry and thirsty mice ($P > 0.4$, two-tailed Wilcox rank-sum test).

i. Similar activity manifold shape (topological dissimilarity, $\beta 0$ features) across hungry and thirsty mice performing a similar behavioral task with different motivations (hunger and thirst) and different rewards (food and water; $P > 0.25$, two-tailed Wilcox rank-sum test).

j. Comparison of estimated intrinsic dimensionality in InsCtx and cortical visual areas in hungry mice performing the same behavioral task during the same motivational state. Intrinsic dimensionality was significantly higher in visual areas ($P < 0.015$, one-tailed Wilcoxon rank-sum test).

k. Comparison of activity manifold shape (topological dissimilarity, $\beta 0$ features) between InsCtx and visual areas in hungry mice performing the same behavioral task. The shape of the activity manifold was significantly more different, i.e., less consistent, in visual areas ($P < 0.0035$, one-tailed Wilcox rank-sum test).

962 **Figure 2**



963

964 **Figure 2 – InsCtx activity dynamics within the manifold are stereotyped across**
965 **different mice**

966 a. Example visualization of time-points on the activity manifold that are associated with
967 certain behavioral variables: red points depict time points during denoted trial types.
968 Different colors in the clustered structure reflect the assignment of different timepoints
969 into neuronal clusters based on the automated clustering method (see Supp. Fig 3). Each
970 row depicts a different plane of the activity manifold.

971 b. Mutual Information between behavioral variables and the activity manifold. Pupil size,
972 cumulative consumed rewards, and rewarded trials had significantly greater mutual
973 information compared to other variables ($P < 0.0004$, one-tailed Wilcoxon signed-rank
974 test with Bonferroni correction). Cumulative consumed rewards had significantly more
975 mutual information than rewarded trials ($P < 0.02$, one-tailed Wilcoxon signed-rank test
976 with Bonferroni correction).

977 c. The most informative variables in 'b' exceeded chance levels, assessed by
978 comparisons to randomly shuffled labels ($P < 0.00095$, one-tailed Wilcoxon signed-rank
979 test).

980 d. Example cluster sequence trial dynamics in two different mice. Left: two representative
981 planes of the clustered structure (mouse 1 is the same as in 'a'), with different clusters
982 represented distinctly by different colors. Right: example of the cluster sequence
983 dynamics for different trial types and outcomes (same cluster colors as the structures on
984 the left). Note that for each mouse, cluster indices (colors) are randomly selected during
985 the clustering procedure. Rewarded trials exhibit a consistent sequence of cluster
986 transitions, while dynamics in other trial types appear less consistent.

987 e. Distribution of trial similarity (cosine similarity) within each different trial type using data
988 pooled from 14 datasets from 5 mice. The distribution of rewarded trials had the highest
989 cosine similarity (0.32 ± 0.37; mean ± SD) and was significantly higher than all other trial
990 types ($P < 5 * 10^{-10}$, two-sample Kolmogorov-Smirnov test with Bonferroni correction).

991 f. Decoding procedure across datasets. The cluster labels of different timepoints are
992 translated across datasets by aligning the order of the transition probability matrix

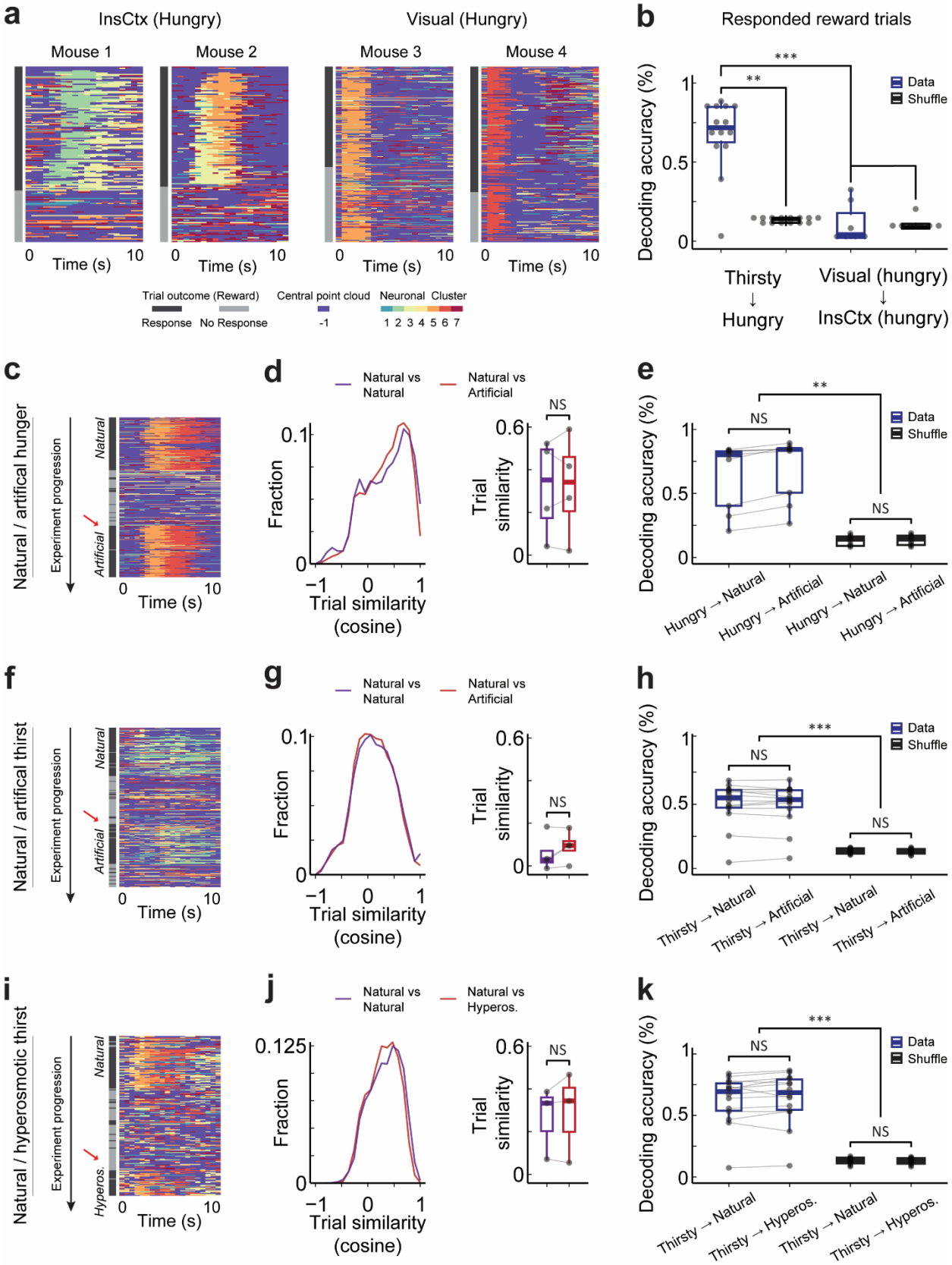
993 between the different neuronal clusters. The translated dataset is then used for training
994 to decode the different trial outcomes of the test dataset.

995 g. Decoding accuracy for different trial types. We used 14 datasets for decoding across
996 mice and 12 datasets for decoding across days. Rewarded trials were decoded
997 significantly better both across days and across mice compared to chance level. Notably,
998 decoding accuracy across days and across mice did not significantly differ ($P < 0.00075$
999 for across days and across mice versus shuffle, one-tailed Wilcox signed-rank test with
1000 Bonferroni correction; $P > 0.2$ for comparison between across days and across mice, two-
1001 tailed Wilcox rank-sum test with Bonferroni correction). Other trial types were not decoded
1002 better than chance levels, whether decoding across days or across mice ($P > 0.2$, one-
1003 tailed Wilcox signed-rank test with Bonferroni correction).

1004 a-g. N=14 datasets from 5 mice

1005

1006 **Figure 3**



1007

1008

1009 **Figure 3 – InsCtx activity dynamics within the manifold are stereotyped across**
1010 **different motivational states and rewards**

1011 a. Visualization of reward cue trial dynamics in two representative mice from InsCtx and
1012 visual areas from hungry mice. Notably, activity in InsCtx displays a consistent and
1013 stereotyped pattern of activity in hungry mice during reward cue trials that is not apparent
1014 after satiation, similar to that of thirsty mice (Fig. 2). In contrast, activity in visual areas
1015 primarily exhibits activity during the presentation of the visual cue and is not affected by
1016 whether trials were rewarded or not (black and grey vertical bars, respectively).

1017 b. Decoding accuracy of rewarded trials from hungry mice. The decoder was trained using
1018 datasets from InsCtx in thirsty mice or from visual areas in hungry mice. Decoding
1019 accuracy was significantly higher when training the decoder on InsCtx of thirsty mice as
1020 compared to visual areas of hungry mice ($P < 0.0002$ for decoded hungry data trained on
1021 thirsty compared to decoded hungry trained on visual area; $P > 0.35$ for decoded hungry
1022 data trained on visual areas compared to shuffled data; $P < 0.0065$ for decoded hungry
1023 data trained on thirsty compared to shuffled data, Dunn test with Bonferroni correction).
1024 N=9 datasets from 6 mice.

1025 c. Left: Example activity dynamics of reward cue trials. The datasets were obtained before
1026 and after artificial induction of hunger using chemogenetic activation of AgRP "hunger
1027 neurons". The red arrow indicates the first responded reward cue trial following the
1028 reinstatement of behavior due to the artificial induction of the motivational state. Note that
1029 the neuronal dynamics of rewarded trials before and after activation are similar.

1030 d. Distribution of pairwise trial similarity (cosine similarity) among rewarded trials across
1031 natural and artificially induced hunger. Left: the distributions pooled from all datasets.
1032 Right: paired comparisons of the mean pairwise trial similarity for each dataset trials
1033 across natural and artificially induced motivational states. The artificially induced
1034 motivational states did not significantly alter the dynamics ($P > 0.15$, two-tailed paired t-
1035 test).

1036 e. Comparison of the decoding accuracy of rewarded trials across natural and artificially
1037 induction of hunger. The artificially induced motivational states did not impact the

decoding accuracy of responded reward trials, which remained significantly higher than chance ($P > 0.3$, for comparisons between natural and artificial data across all states; $P > 0.08$ for comparisons between natural and artificial shuffles across all states; $P < 0.008$ for comparisons between data and shuffle for both natural and artificial; one-tailed Wilcox signed-rank test with Bonferroni correction).

f-h. Same as in ‘c’-‘e’ before and after artificial induction of thirst using chemogenetic activation of SFO “thirst neurons”.

i-k. Same as in ‘c’-‘e’ with before and after induction of hypertonic thirst using injection of hypertonic saline.

a-k. Data from N=10 datasets (N=7 mice) for hungry datasets, N=14 datasets (N=5 mice) for thirsty datasets, and N=7 datasets (N=5 mice) for visual areas datasets, N=4 datasets (4 mice) for AgRP activation, N=4 datasets (N=4 mice) for SFO activation, N=3 datasets (N=3) mice for hypertonic saline injection.

1052 **Figure 4**

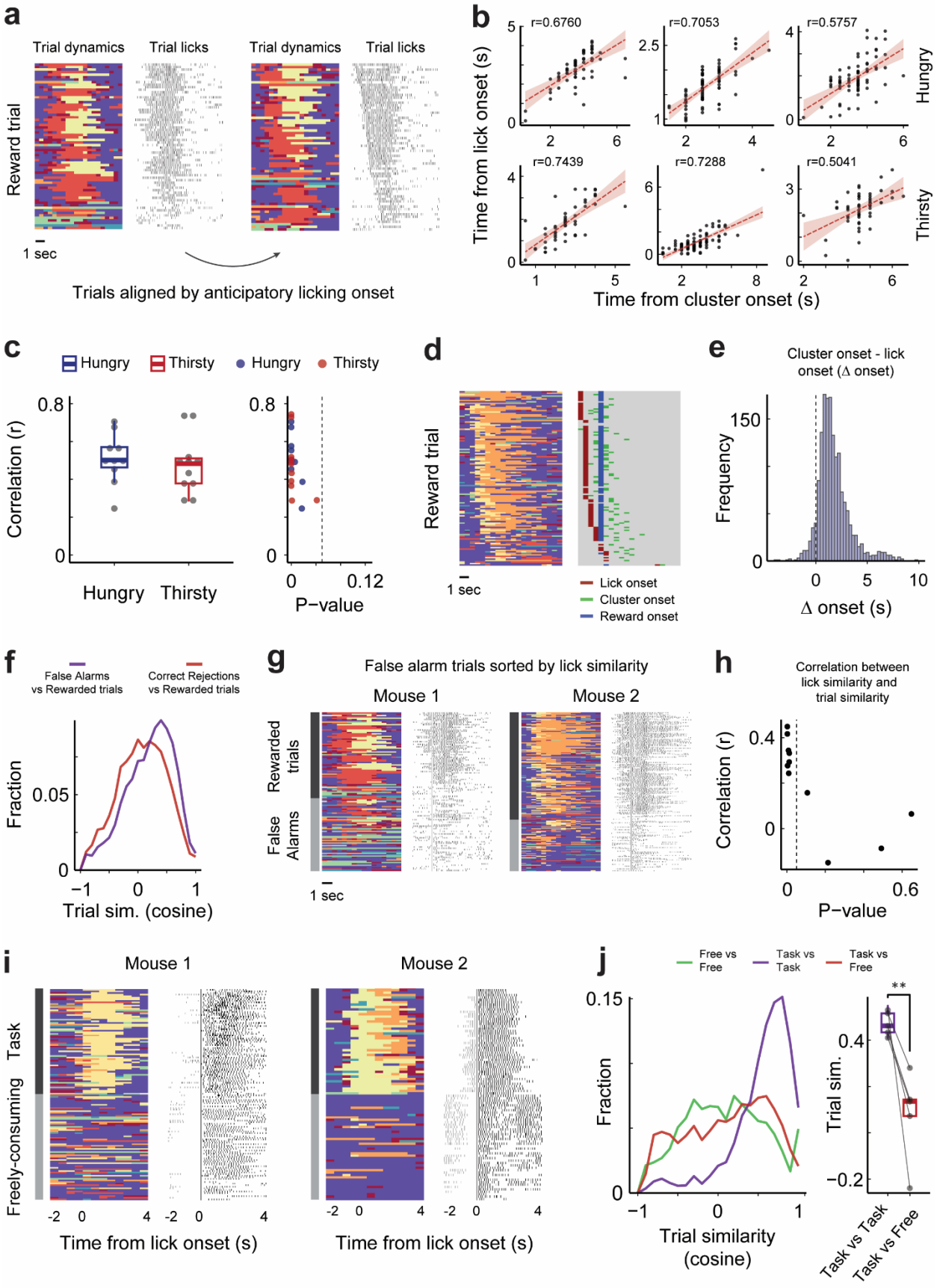


Figure 4 – Stereotyped activity dynamics reflect goal-directed reward anticipation, but not licking, taste, or valence

a. Example of neuronal dynamics during rewarded trials and task-related licking behavior. The onsets of sequential cluster patterns appear to be synchronized with the onset of anticipatory licking.

b. Examples of correlations between anticipatory licking onsets and rewarded trials cluster onsets in 3 datasets from hungry mice (top) and 3 datasets from thirsty mice (bottom). Note relatively high positive correlations in all examples.

c. Left: Distribution of all correlations across all hungry and thirsty datasets (N=10 and N=11, respectively). Right: Scatter plot describing correlation levels as a function of significance levels. All datasets exhibit significant positive correlations between anticipatory licking onset and the rewarded trial sequential cluster pattern onset, as indicated by the dashed vertical line at P=0.05 ($P < 0.04$, for all datasets, Pearson's product-moment correlation).

d. Example visualization of reward onset, anticipatory licking onset, and activity cluster onset. Left: Cluster dynamics of all rewarded trials. Right: Onsets of anticipatory licking, reward delivery, and cluster dynamics per trial. Note that the onset of the sequential cluster pattern consistently follows the onset of anticipatory licking and appears unrelated to the onset of reward delivery.

e. Distribution of the time difference between rewarded trial cluster onset and anticipatory licking onset. Across all trials, cluster onset occurs approximately 1 second *after* lick onset.

f. Distribution of trial similarity between incorrectly responded aversive and neutral trials ("false alarms") and rewarded trials (purple), compared to the similarity between correctly non-responded aversive and neutral trials ("correct rejections") to rewarded trials (red). Notably, false alarm trials are more similar to rewarded trials compared to correct rejection trials. (Data from N=14 datasets from N=5 thirsty mice).

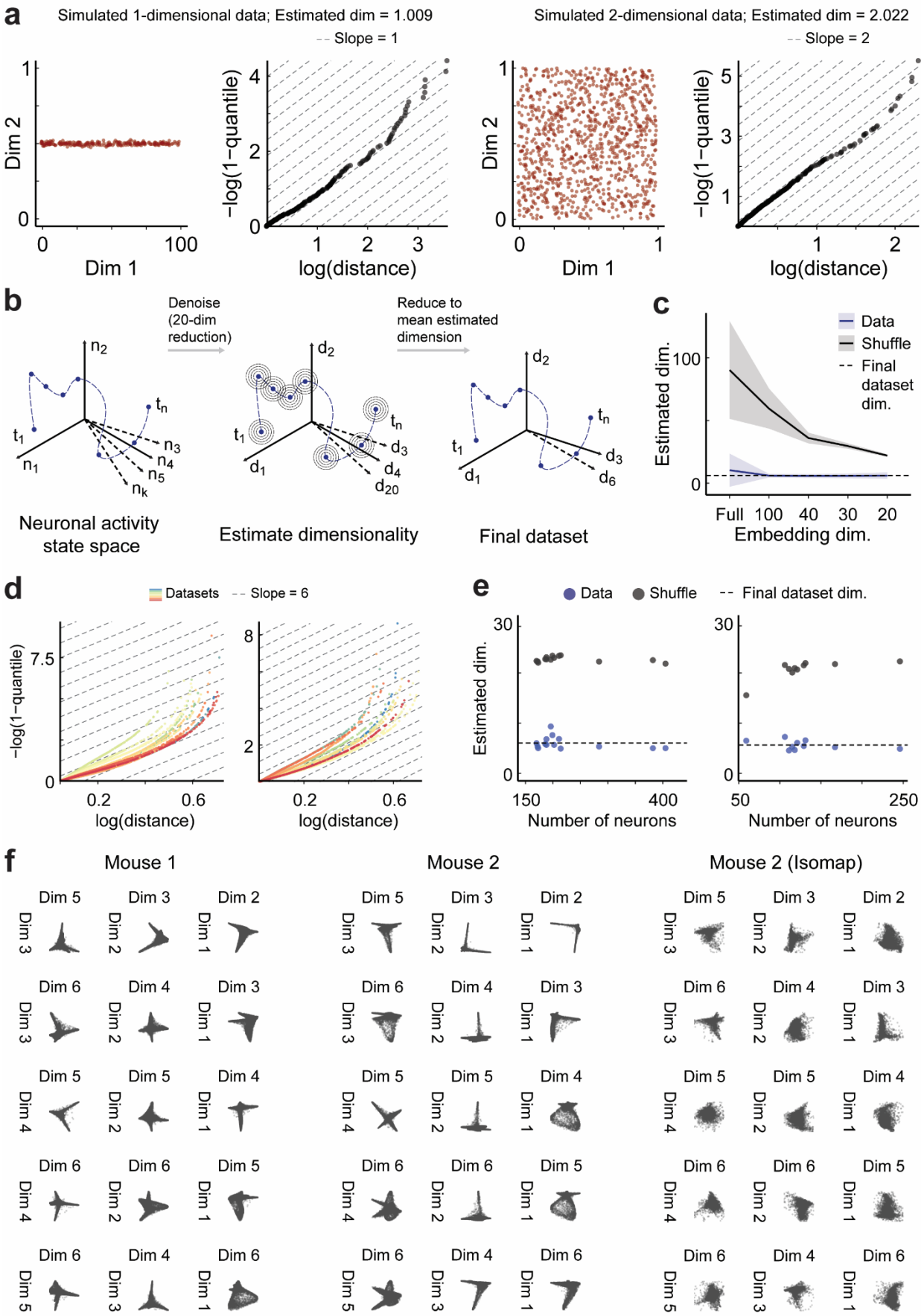
g. Two examples of cluster sequence trial dynamics during false alarm trials and rewarded trials. False alarm trials were sorted based on their *licking* pattern similarity to the *licking* patterns in rewarded trials. Note that as the licking pattern becomes more similar, false alarm cluster sequence trials become more similar to rewarded trials.

h. Distribution of all correlations between licking similarity and trial similarity in false alarm trials and rewarded trials, as a function of significance levels. Out of 11 datasets, 9 had positive correlations, and 7 of these correlations were statistically significant ($P < 0.05$, assessed using Pearson's product-moment correlation).

i. Example cluster sequence dynamics and licking behavior in two mice during reward consumption while engaged in the operant task (rewarded trials) and during self-paced free consumption behavior. The dynamics were aligned based on the onset of licking. The neuronal cluster dynamics during operant goal-directed behavior appear distinct from those observed when the mice are freely consuming rewards.

j. Left: Distribution of all pairwise comparisons between reward consumption trials during task engagement and free consumption behavior (pooled from N=5 datasets from N=5 mice). Freely consuming mice exhibit distinct cluster dynamic, reflected by the distribution of task versus free behavior trial similarity centered around zero. Right: Comparison of the average trial similarity for each dataset individually when comparing task dynamics to itself and task dynamics versus freely behaving consumption dynamics. Free consumption dynamics are significantly different than task consumption dynamics ($P < 0.009$, one-tailed paired t-test)

1103 **Supplemental Figure 1**



Supp. Figure S1 – Dimensionality estimation and reduction

a. Dimensionality assessment using a two-NN Estimator[42] on two simulated example datasets. Left: Estimation of a simulated one-dimensional dataset. Right: Estimation of a simulated two-dimensional dataset. In both datasets dimensionality is accurately approximated, with data points aligning parallel to the simulated dimension (i.e., slope).

b. Datasets initially undergo an intermediate dimensionality reduction, followed by an estimation of dimensionality for the intermediate dataset. The final dataset is achieved after an additional iteration of dimensionality reduction, using the mean estimated dimension in the intermediate iteration across all datasets.

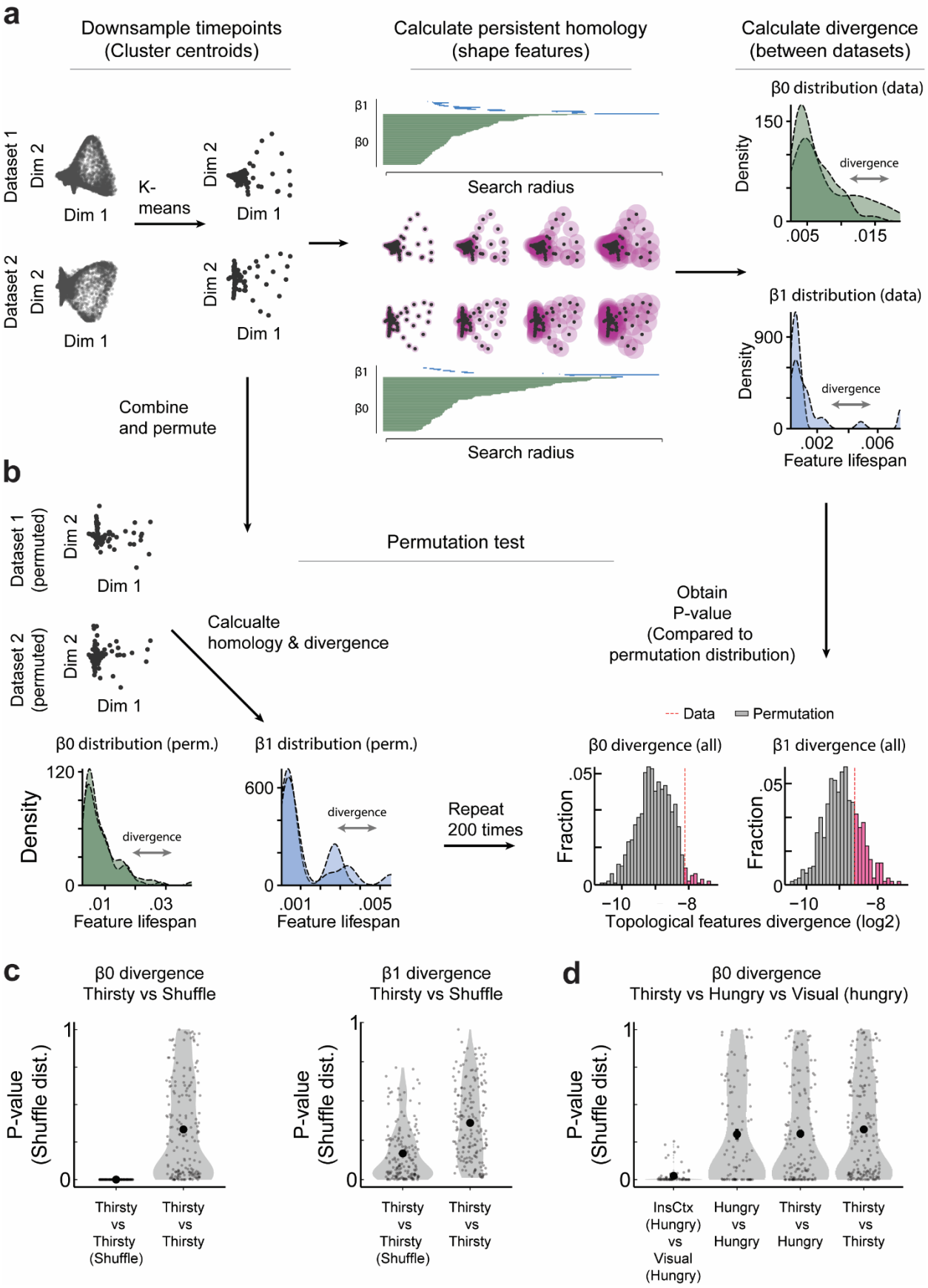
c. Estimation of intrinsic dimensionality in shuffled and original data across different intermediate dimensionality reductions and the full embedded data with no intermediate dimensionality reduction. Estimation of the intrinsic dimensionality in the original data converges to ~6 (dashed line, chosen for subsequent dimensionality reduction, see Supp. Figure 2). In contrast, estimation on shuffled data yields an intrinsic dimension close to the intermediate embedding dimension.

d. Left: Estimated dimensionality for 14 different datasets of thirsty mice. The near-constant slope observed across all datasets suggests a consistent intrinsic dimensionality of 6, which suggests the presence of approximately six latent variables within the activity manifold. Right: Same as in `left` for 10 different datasets of hungry mice.

e. The estimated dimension was not correlated with the number of neurons sampled in each dataset for real data and shuffled data. Left and Right are the same as in `d` ($P > 0.2$, Pearson's product-moment correlation).

f. All dimension pairs in two different mice. Mouse 2 is presented with two different non-linear dimensionality reduction algorithms (right: Isomap; left: Laplacian eigenmaps, used throughout our analyses).

Supplemental Figure 2



Supp. Figure S2 - Assessing similarity between datasets using topological data analysis

a. Illustration of the procedure used to assess the similarity between pairs of datasets. Initially, timepoints are down-sampled through K-means clustering. Subsequently, the distribution of topological features is extracted independently for each dataset, The magnitude of divergence in the feature distributions directly reflects the dissimilarity between the datasets, providing a quantitative measure of their dissimilarity.

b. Description the process for conducting permutation tests to assess the statistical significance of dissimilarity between datasets. Datasets are combined and permuted and redivided 200 times. During each permutation, the divergence of topological features between the randomly permuted and divided datasets is calculated. Subsequently, a p-value is derived by comparing the divergence observed between the original datasets to the distribution of divergences constructed from the permutations. This allows for the assessment of the statistical significance of dataset dissimilarity.

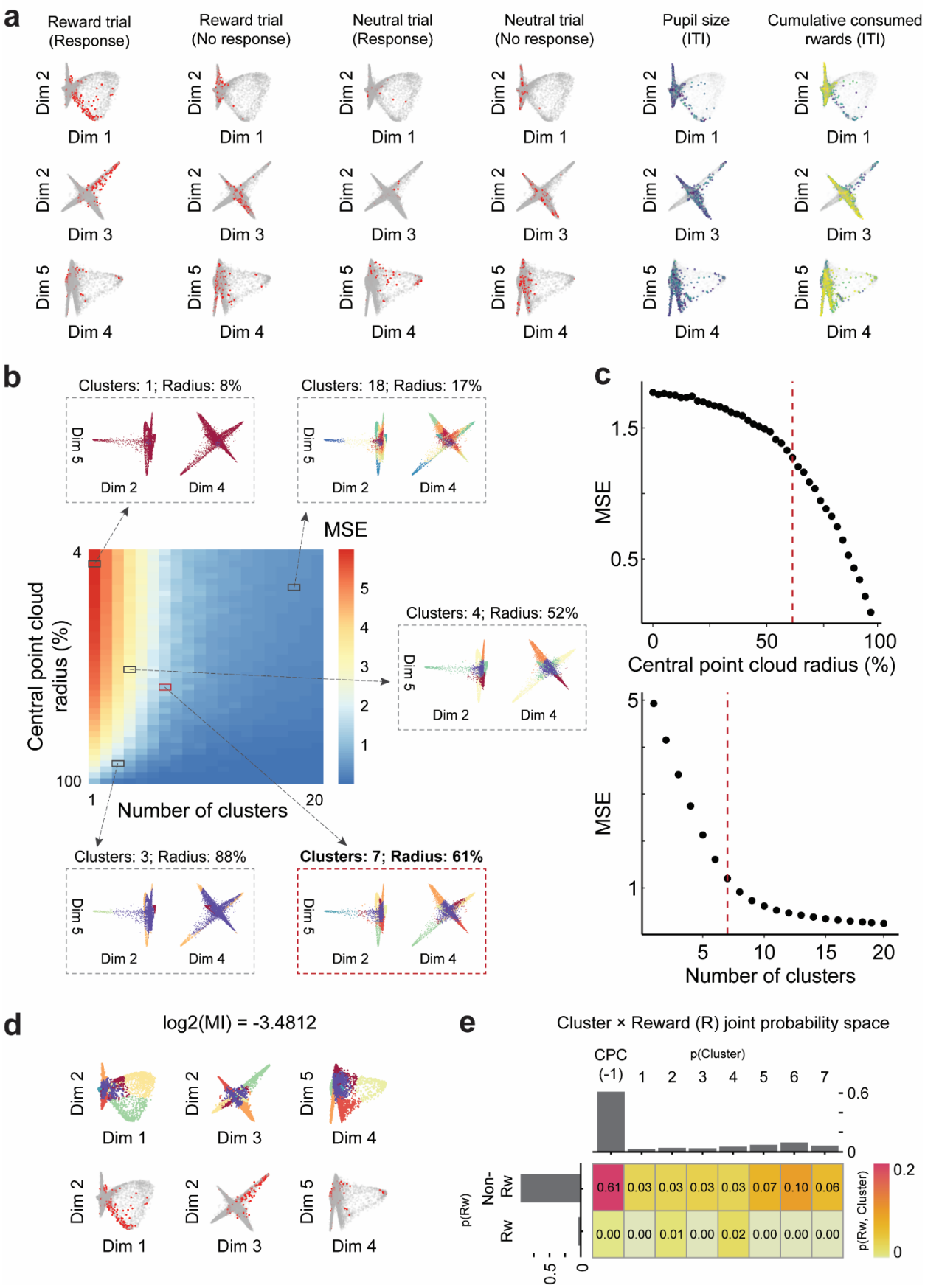
c. P-values calculated based on the shuffle distribution for all pairwise comparisons assessing the topological divergence of β0 features.

d. Same as in `c` but for $\beta 1$ features. Shuffle distributions compared to neuronal data had non-significant different results. This observation may be attributed to the sparse nature of the β1 features extracted from each dataset.

e. Same as in `c` but also for InsCtx and visual cortex in hungry mice.

c-e. N=100 pairwise comparisons for hungry vs hungry, N=182 pairwise comparisons for thirsty vs thirsty, N=140 pairwise comparisons for thirsty vs hungry mice, N=70 comparisons for Visual vs InsCtx mice, from N=14 datasets of thirsty mice, N=10 datasets of hungry mice, N=7 datasets of visual mice.

Supplemental Figure 3



Supp. Figure S3 – Automated clustering of activity manifold; clustered structure relationship with experimental variables

a. Depicting the activity manifold annotated with external/internal variables in an exemple dataset. In each row, a distinct plane of the activity manifold is presented. To illustrate discrete variables (columns 1-4; responded reward trials to non-responded neutral trials) we use red points to indicate the event occurrence. In the case of continuous variables (columns 5-6; pupil size and cumulative consumed rewards) points are color-coded, ranging from blue to yellow according to the variable's values.

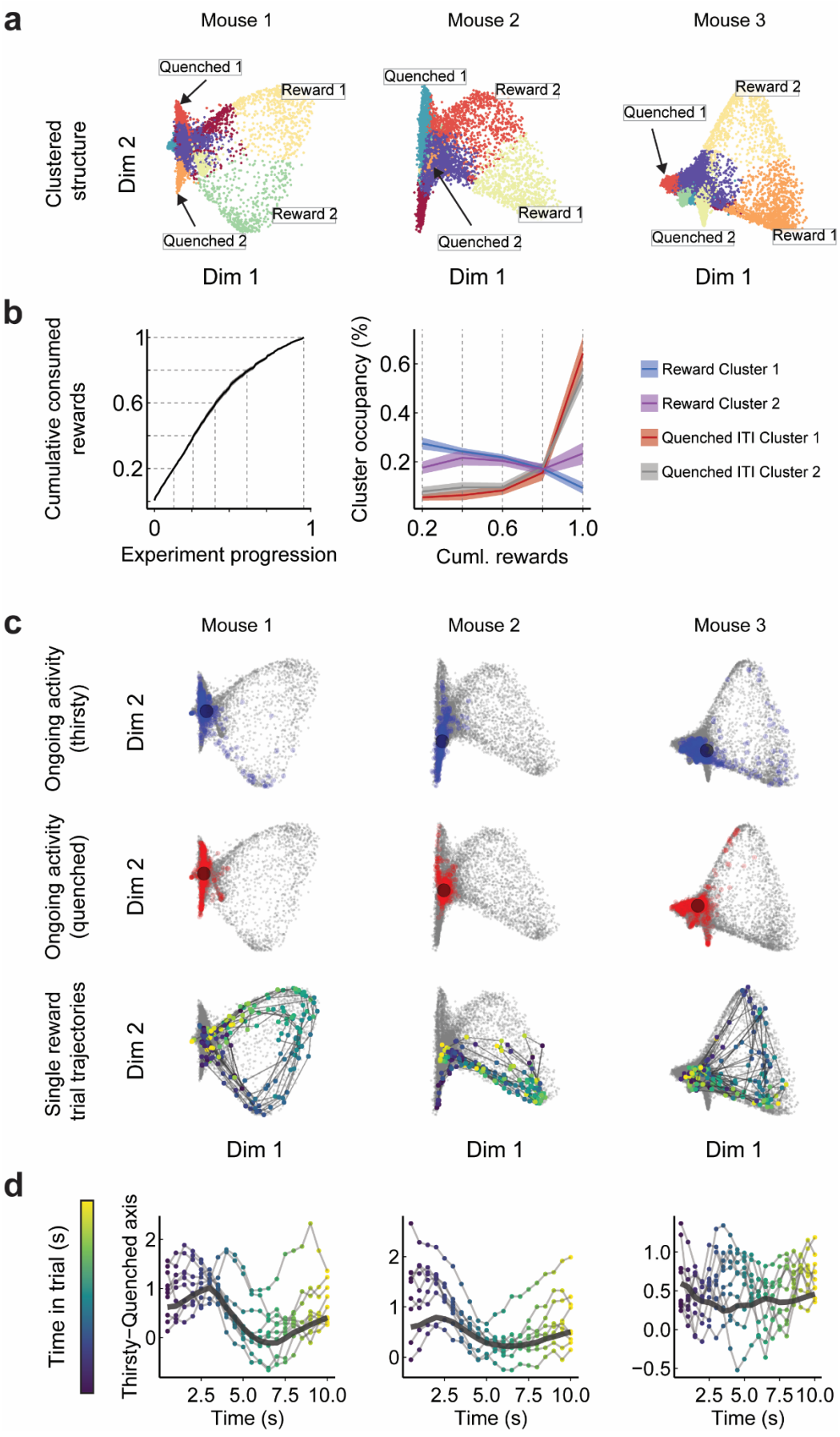
b. Illustration of the automated clustering process. For every dataset, we construct a configuration matrix, as depicted. This matrix is generated by performing K-means clustering while iteratively adding points into the central point cloud (rows) and testing with varying cluster numbers (columns). The Mean Squared Error (MSE) is calculated for each K-means configuration during each repetition and stored in the corresponding matrix cell. Several examples of different configurations are outlined with dashed gray lines, whereas the final selected configuration is delineated with a dashed red line (see 'c' for details).

c. Identification of the optimal configuration. The optimal number of points to include in the central point cloud and the optimal cluster count are automatically determined by identifying the deflection point, as detailed in the methods section.

d. Example planes of the clustered structure alongside an external variable of interest

e. Illustration of calculation of Mutual Information. The Mutual Information between the clustered structure and the external variable is obtained by calculating their joint probability space.

1184 **Supplemental Figure 4**



Supp. Figure S4 – Comparison of linear projection dynamics with activity manifold dynamics

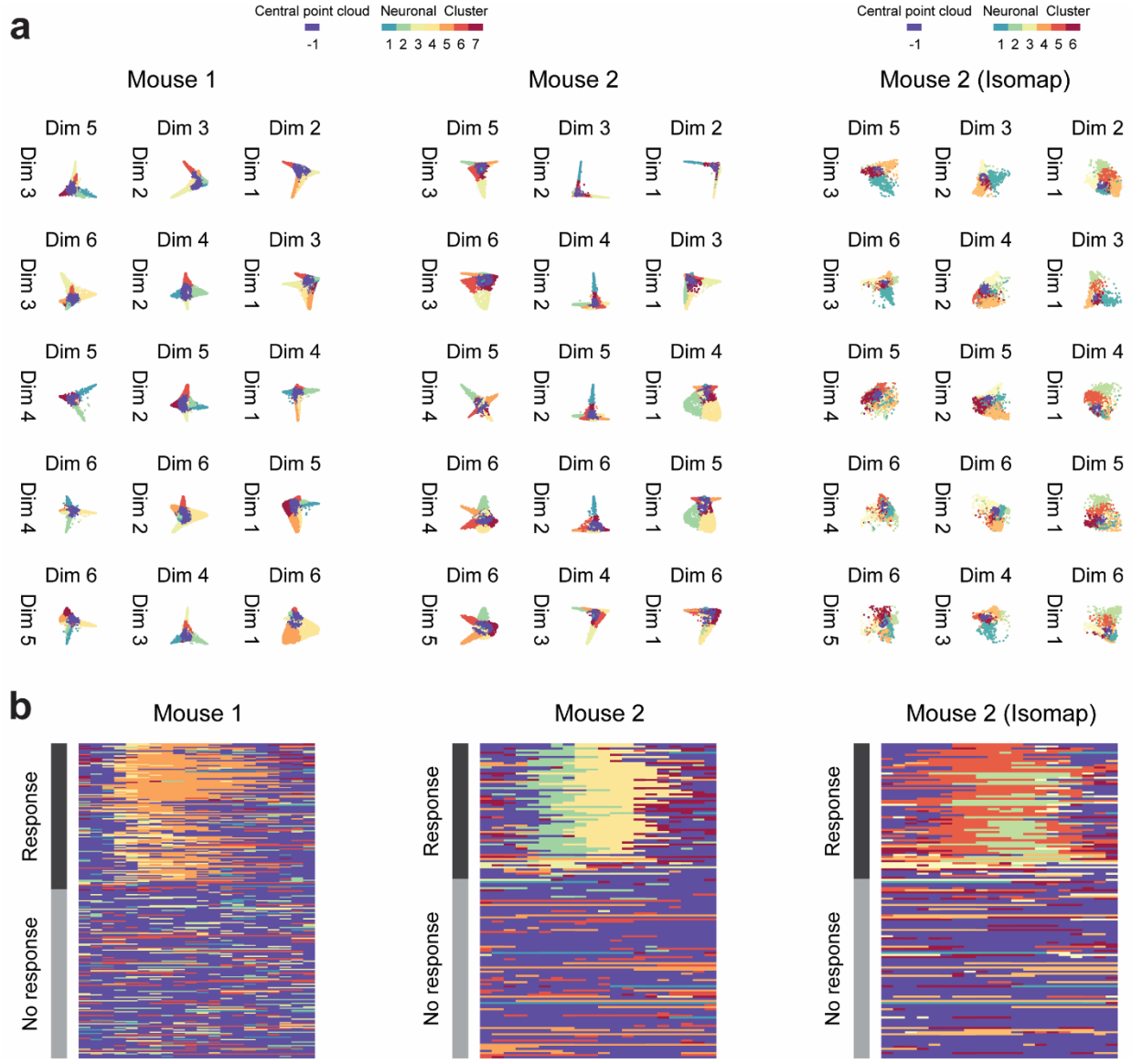
a. Activity manifold from three example mice. Top: Neuronal clusters annotated using automated clustering, with labeled clusters representing ITI water satiety and reward clusters used in ‘b’.

b. Left: Physiological state approximated by cumulative consumed water rewards over time (mean across all datasets). Dashed lines indicate the experiment time at which mice reached a certain quenched level (e.g., by the middle of the experiment, mice consumed 70% of the rewards they would consume). Right: neuronal cluster occupancy (percent of time that activity occupied a certain cluster) during reward epochs and quenched phases in relation to gradual changes in physiological state (corresponding to dashed lines in the left cumulative rewards plot). Occupancy of reward clusters remained relatively consistent throughout gradually changing physiological states, and slightly decreased as the number of responded reward trials decreased with water satiation. In contrast, the two most frequently occupied neuronal clusters during the quenched phase were primarily visited when animals ceased water consumption and became quenched, but not before.

c. Top, middle: Timepoints of ongoing activity used to calculate average activity vectors for thirst (blue) and quenched (red) states, indicated by black centroids (following our previously published procedures[33]). Bottom: Trajectory of ten reward trials projected on the activity manifold.

d. The same ten reward trials as in ‘c’, projected linearly using a neural mode obtained from subtracting the average ongoing activity in thirst and quenched states[33]. During reward trials, population activity briefly shifts from a thirst state towards a quenched state. The trajectory of the same trials can also be observed with respect to the activity manifold described in ‘c’.

Supplemental Figure 5

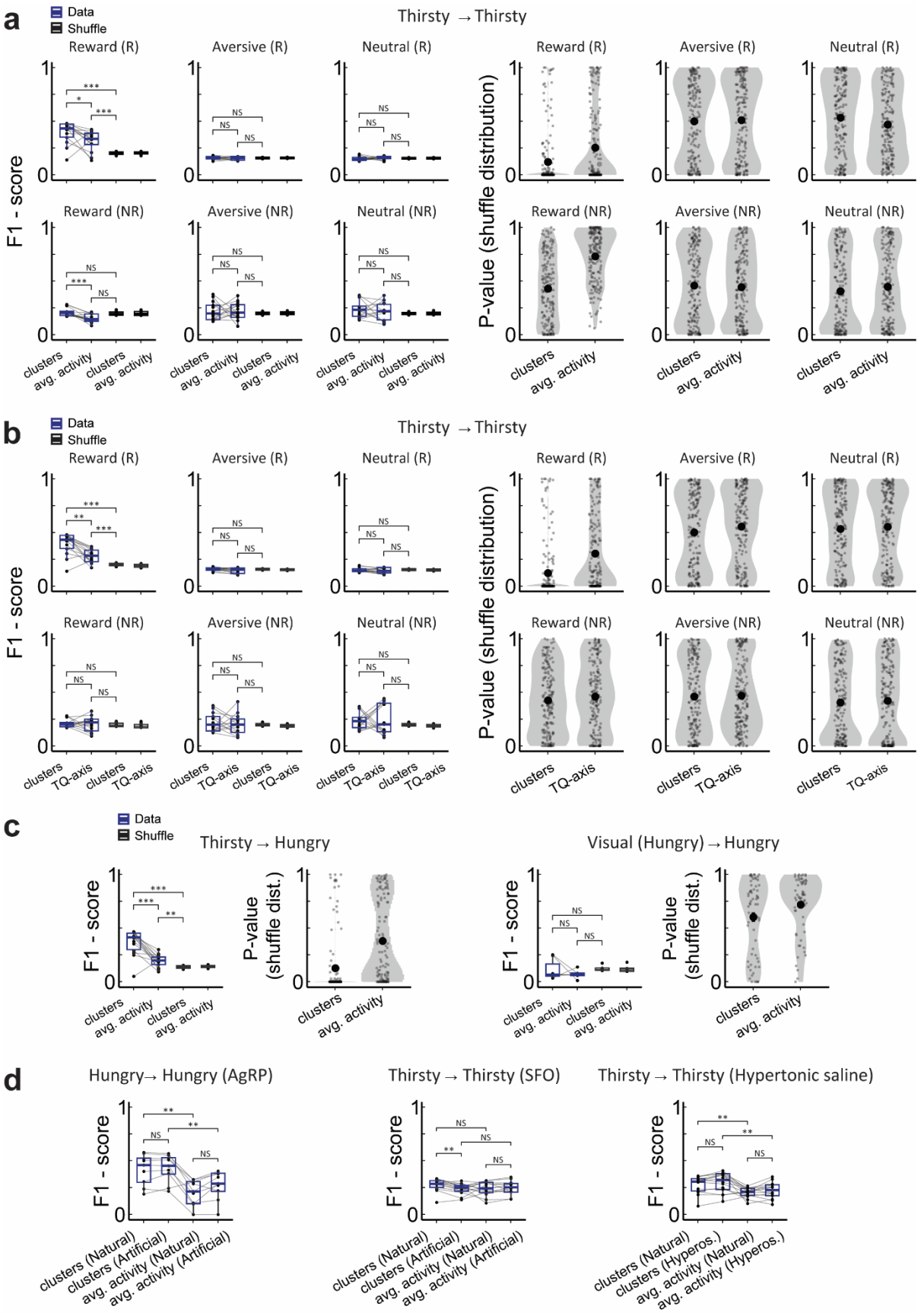


Supp. Figure S5 – Activity dynamics of manifold cluster sequences with two different dimensionality reduction methods

a. All dimension pairs in two different mice annotated by the automated clustering method. Mouse 2 is presented with two different non-linear dimensionality reduction algorithms (Right: Isomap; Left: : Laplacian eigenmaps, used throughout our analyses).

b. Activity dynamics of the same examples as in `a` for reward cue trials. Notice that the stereotyped activity patterns are also observed when using a different non-linear dimensionality reduction algorithm (Isomap).

1224 **Supplemental Figure 6**



Supp. Figure S6 – Analysis of across dataset decoder performance.

Left: Assessing decoder F1-scores for trial types. A comparison of F1-scores when employing a decoder trained on manifold cluster sequences and another on average activity levels. Notably, only rewarded trials exhibited significantly higher decoding performance above chance. Additionally, decoding using manifold clusters outperformed decoding based on average activity levels. Right: The distribution of P-values for decoding derived from the surrogate shuffle distribution for each pairwise decoding is shown. All trial types exhibited uniform p-values in terms of F1-score significance levels, except for rewarded trials indicating significant decoding.

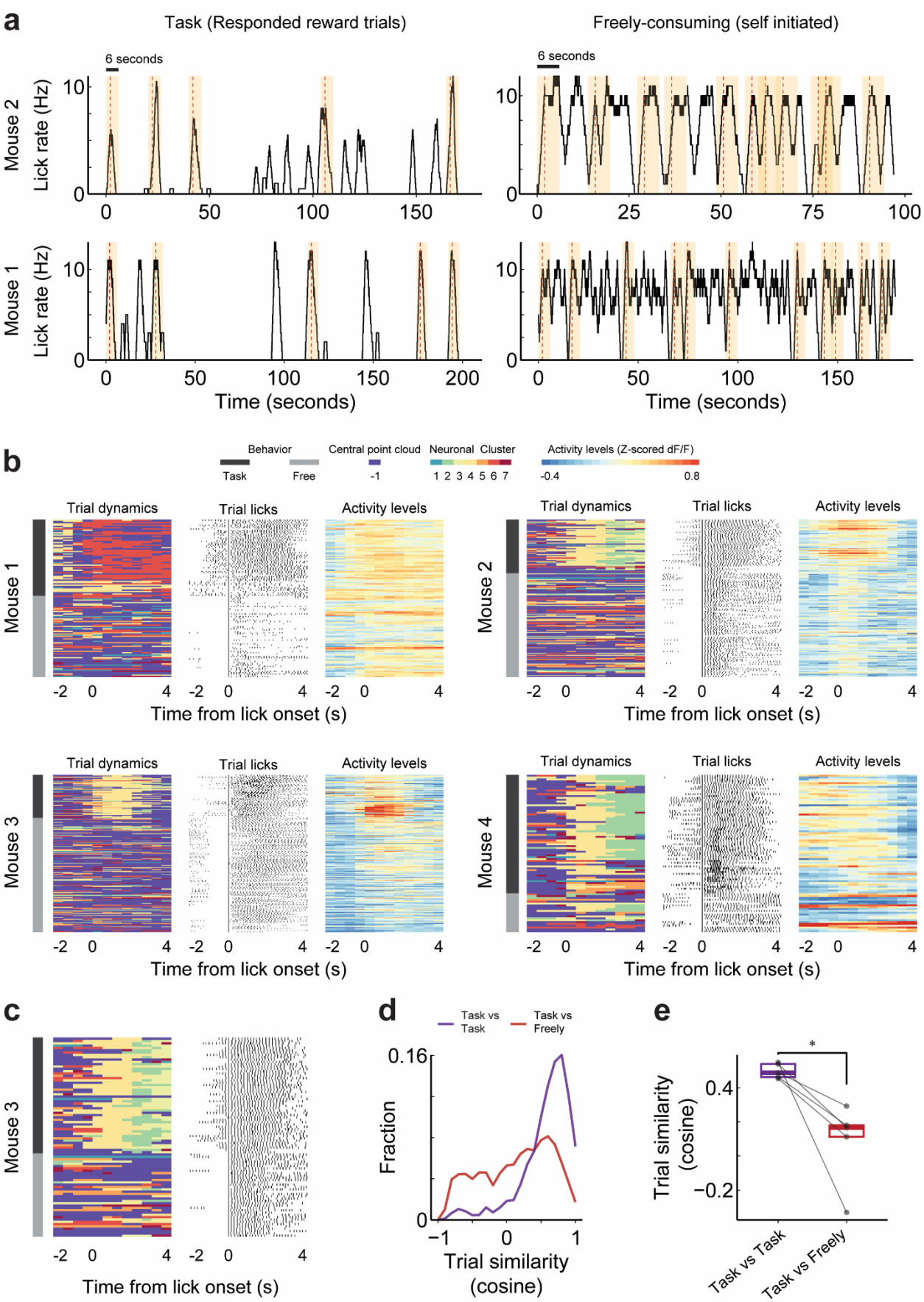
b. Same as in `a` but now comparing the performance of a decoder trained on manifold cluster sequences versus one trained on the projection of trials onto thirsty-quenched axes (T-Q axis).

c. Same as in `a`, but now comparing between a decoder trained on InsCtx thirsty or visual cortex hungry datasets, and tested on InsCtx hungry datasets accordingly.

d. Comparison of the F1-score for rewarded trials across natural and artificial motivational states.

a-d. N=14 thirsty datasets, N=10 hungry datasets, N=7 visual datasets, N=4 AgRP datasets, N=4 SFO datasets, N=3 hyperosmotic datasets. (Significance levels: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, one-tailed Wilcoxon signed-rank test).

1245 **Supplemental Figure 7**



Supp. Figure S7 - Comparison of activity dynamics in goal-directed behavioral task and free consumption of rewards

a. Examples of the pseudo-trial construction process to compare reward consumption behavior in two mice. Left: In the visual discrimination task, trials were generated by analyzing the 2 seconds before and the 4 seconds following lick onset in rewarded trials. Right: In free reward consumption, trials were generated by analyzing the 2 seconds before and 4 seconds after self-initiated lick bouts (timepoints in which licking rate sharply increased after a sufficient decrease). Only non-overlapping trials were chosen.

b. Example of activity dynamics during goal-directed behavior (operant visual discrimination task) and free reward consumption (self-initiated) in four different mice. Heatmaps are aligned with the constructed lick trials as described above. Left: Sequence of cluster dynamics observed in constructed lick trials. Middle: Licking in the constructed lick trials. Right: Overall mean activity levels. While goal-directed and self-initiated trials exhibit distinct dynamics, there is an increase in overall activity levels during both goal-directed and self-initiated consumption of rewards.

c. Example cluster sequence dynamics and licking behavior, while controlling for variations in licking behavior between task-based consumption and freely initiated consumption. We addressed these behavioral distinctions by specifically choosing free-consumption trials in which the licking rate within the initial two seconds of the licking onset fell within the mean and standard deviation of the licking rate within the first two seconds of task-based lick bouts.

d. Distribution of all pairwise trial dynamics comparisons between reward consumption trials during task engagement and free consumption behavior, when controlling for behavioral differences in licking (pooled from N=5 datasets from N=5 mice).

e. Comparison of the average trial similarity for each dataset individually when comparing task dynamics to task dynamics, and task dynamics versus freely behaving consumption dynamics, while controlling for behavioral differences in consumption ($P < 0.05$, one-tailed paired t-test)