Sparse representations of sensory signals in neuronal networks

Yael Ben Nahum[1], Noa Elnhorn[2], Hadas Benisty[3]

[1]Faculty of Biomedical Engineering, Technion – IIT, Haifa, Israel

[2] Faculty of Electrical & Computer Engineering, Technion – IIT, Haifa, Israel

[3] Faculty of Medicine, Technion – IIT, Haifa, Israel

Abstract: Learning is a contentious process in which the brain adaptively changes its activity to improve performance. Outcome representation by cortical networks during learning has been the focus of many research groups in the past decade, where most work has been done regarding binary outcome – success or failure in performing a task. Still, it is plausible to assume that neuronal networks use a richer representation and thus report a continuous value of outcome to achieve optimal performance faster. To test this hypothesis, we analyzed the neuronal activity of Pyramidal Cells in layer 2-3 of the primary motor cortex (M2), recorded from mice performing a hand-reach task. In this setup, mice are trained to reach grab and consume a (plain) food pellet. Once mice are trained, they are introduced with flavored pellets - sweet and bitter, along with the plain ones used for training. Here we introduce a multivariate analysis pipeline for the detection of sub-populations encoding binary outcome (success/failure) and value of outcome (flavor). Our results confirm the existence of a sub-population of cells reporting binary outcomes well after movement is over. Once flavors are introduced, the encoding of binary outcome changes but does not vanish. Representation of flavor emerges with exposure to flavor, where at first novelty is reported and then the value of the outcome (aversive vs. tasty) is reported. Interestingly, flavor is also reported during the preparatory time segment, before the go-cue. Overall, our results confirm the hypothesis of a rich representation for outcome, not only as a binary success/failure value after movement is complete, but as a continuous variable reported during the whole trial period depending on the experimental context.

1. Introduction

The primary motor cortex (M1) is crucial for motor performance. To make ongoing adjustments and plan future movements, the central nervous system monitors action outcomes. Action outcomes can be evaluated either as a binary variable, i.e., as a successful or failed attempt, or in a non-binary fashion, as a continuous value.

Previous work shows that binary outcome (success or failure) is reported by the activity of a small sub-population of layer 2-3 pyramidal neurons (PNs) in the M1 cortex of mice [1], where this representation gradually emerges along the training process, rather than innate [2].

Here we aim to explore if and how the value of outcome is reported as well. To this end, we use 2-photon Calcium imaging data of PNs from layer 2-3 of M1, collected from head-fixed mice performing a hand-reach task of grabbing and consuming a food pellet (similar to [1] and [2]). The experimental team used flavored pellets representing different values – sweet (sucrose) and bitter (quinine), see Fig. 1.

Previous work showed that in this experimental setup, the encoding of flavor as a continuous outcome variable emerges with exposure to flavor, at the single-cell level [3]. Here we set out to understand the relationship between the encoding of binary outcome (success/failure) and the encoding of reward-value (flavor) at the population level. Specifically, we would answer the following questions: 1) How does exposure to different values affect (binary) outcome signaling? 2) Does the network report value as well? 3) What is the relationship between binary and value signaling of outcome?

Sparse Principal Component Analysis (SPCA) was used by Lin et al. [4] to explore the representation of sensory data and memory at the population level of the prefrontal cortex in monkeys during a working memory task. As opposed to Principal Component Analysis (PCA) which maximizes the explained variance of the signal, SPCA is more efficient and biologically plausible, considering the mismatch between the relatively low number of possible connections and the high number of potentially informative neurons [4]. In this work, we use SPCA to extract a low-dimensional representation of the population dynamics. Then we use a set of classifiers to understand the information encoded within each component. We apply this strategy to different experimental conditions which allows us to track the emergence of flavor encoding, its relationship with the encoding of binary outcome, and how this affects the network's structure.

Overall, our findings present a small but significant contribution to the growing body of knowledge of the way outcome and reward signals are represented in M1. Understanding these mechanisms could progress the field toward a better understanding of motor learning in healthy conditions as well as degenerative ones.

2. Methods

2.1. Experimental apparatus

In our project, we used imaging data collected by Zohar Lotan from Jacky Shiller's lab in the Technion Faculty of Medicine [3].

Calcium imaging was used to track the activity of layer 2-3 neurons in M1 of mice performing a hand-reach task of grabbing and consuming food pellets positioned on a rotating table (Fig. 1). The task was performed while mice were head-fixed and in dark lighting conditions thus visual cues did not play a role in the performance.

Each mouse participated in multiple experimental **sessions** on different days. An experimental session consisted of 60-120 **trials**. A single trial is 12 seconds long, where after the first 4 seconds a tone is played, signaling that the plate is rotating. Then, the food pellet is placed in front of the mouse (Fig. 1D). In our analysis, we define the tone time as 0[s] and segmented trials into four **time-segments**: 'preparatory' from start time to tone minus 1sec, 'movement' from tone minus 1 to tone +2sec, 'consumption' from tone+2sec to tone+5sec and 'end' from tone+5sec to tone+8.

Some food pellets were flavored with quinine (aversive flavor) or sucrose (tasty flavor) and some were plain (neutral flavor, familiar to the animals). The flavors were arranged by the experimental session type, which was one of the following:

- **Train** – baseline sessions in which mice are trained to perform the task. In train sessions pellets are plain (Fig. 1A).

- **First** – the first exposure to flavored pellets. Flavors are presented in batches of consecutive pellets of each flavor (Fig. 1B).
- **Ongoing batch** – Mice were familiar will all flavors. Like first exposure sessions, in batch sessions pellets are arranged in batches of consecutive pellets of each flavor (Fig. 1B).
- **Ongoing random** – Mice were familiar will all the flavors. Unlike batch sessions, in random sessions pellets are arranged in random order of flavor (Fig. 1C).

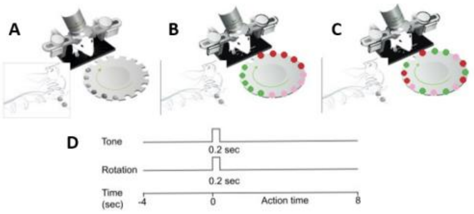


***Figure 1**: Illustration of the experimental setup and trial timeline (adopted from Levy et al. 2020). Mice perform a hand-reach task of grabbing and consuming food pellets positioned on a rotating table. **A**. Train session setting, pellets are plain (neutral flavor, gray). **B**. Batch session setting, pellets are arranged in batches of consecutive pellets of each flavor. **C**. Random session setting, pellets are arranged in random order of flavor. **D**. A trial's timeline.*

2.2. Data analysis methods

Figure 2 presents the analysis pipeline: the analysis input is the neuron activity. First, we applied Sparse Principal Component Analysis (SPCA) to extract a low-dimensional representation of the population dynamics. Second, to evaluate the information encoded in each component under various experimental conditions, the response in each component per time segment is then fed to a set of classification tasks based on Logistic Regression. Third, we then used permutation testing to detect which component encodes the information per time segment. Bookkeeping was performed to store, track, and analyze the neurons and components encoded. The final analyses are based on understanding the fraction of encoding neurons through their contribution to each component. Finally, we examined the emergence of the encoding neurons as the experiments progressed in time and explored the relevance between the importance of sub-populations to encoding and their centrality in the overall network.

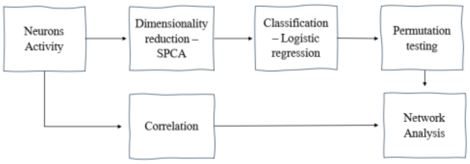

Neurons Activity
Dimensionality reduction – SPCA
Classification – Logistic regression
Permutation testing
Correlation
Network Analysis

***Figure 2**: Analysis pipeline.*

2.2.1. Dimensionality reduction – SPCA

PCA is a dimensionality reduction method that extracts a low dimensional representation of data, based on components that explain the most variance of the data and achieve optimized reconstruction's MSE. PCA optimization equation is:

$$\max v^T \Sigma v \; s.t \; \|v\|_2 = 1$$

where $X \in \mathbb{R}^{S \cdot T \times N}$ is the data matrix and $\Sigma = \frac{1}{S \cdot T} X^T X \in \mathbb{R}^{N \times N}$ is the empirical covariance. SPCA, however, is a constrained version of PCA. i.e., it's also optimized for maximal explained variance but with sparsity constraint:

$$\max v^T \Sigma v$$

$$s.t \; \|v\|_2 = 1, \|v\|_0 \leq MaxL0.$$

We applied SPCA to extract a low-dimensional representation for the dynamics of the neuronal activity. This representation comprises from K **components**, where $K < N$. Feeding the data tensor to SPCA results with a projection tensor of size $K \times S \times T$. Reconstruction is performed similarly to PCA.

SPCA has two parameters: the number of components, K and the maximum number of neurons that can contribute to each component (the L0 norm), $MaxL0$.

The number of components:

Fig. 3 presents the explained variance of SPCA and PCA using a growing number of leading components. The explained variance by PCA is significantly higher since PCA maximizes the variance by construction. SPCA, however, is constrained to sparsity (making it more biologically plausible) at the expense of lower explained variance.

With no apparent trend in the graph to indicate a preferred value of K, we decided on an arbitrary value of 10 components.

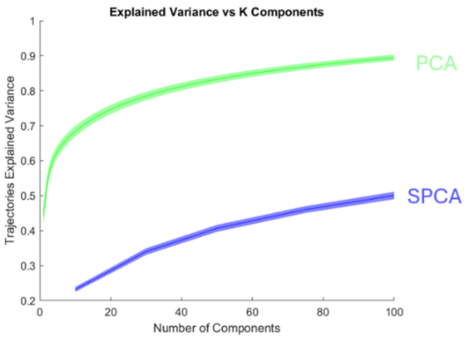


***Figure 3**: Explained variance of reconstructed data as a function of the number of components for PCA and SPCA methods.*

Maximal Number of Cells Per Component:

To choose this parameter we used the next step of the analysis pipeline – classification. We repeated the classification step with different values of $MaxL0$ parameter. For every classifier, we calculated the relative accuracy

$$ACC_{rel} = \frac{ACC - chance}{chance}$$

where $chance = max\left\{\frac{labels==1}{labels}, \frac{labels==0}{labels}\right\}$.

Fig. 4 presents an example of the relative accuracy values for all components (y-axis) and time windows (x-axis) relevant to a specific question and three parameter values. Increasing the number of cells per component increases the number of time windows and components that achieve significant accuracy values. Choosing the parameter was made manually to be small enough to maintain the sparsity and large enough to receive enough information so we could capture the encoding in the classifier. The chosen parameter was $MaxL0 = 30$.

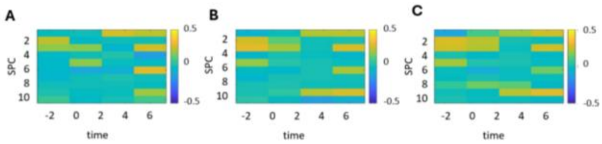


Figure 4: Relative accuracy for all components (y-axis) and time windows (x-axis) 3 different values of $MaxL0$: A. 15, B. 30, C. 45.

2.2.2. Classification – Logistic regression

To evaluate the information encoded in each component under various experimental conditions, we designed a series of **questions** that relate to different characteristics of the outcome.

- **Binary outcome** – Is the outcome encoded and identifiable after the dimension reduction? Relevant separation classes: successful trials vs. failure trials.
- Does flavor affect outcome encoding? Relevant separation classes: successful sucrose trials vs. failure sucrose trials, successful quinine trials vs. failure quinine trials, and successful plain trials vs. failure plain trials.
- **Novelty** – Is familiarity with the flavor encoded? Relevant separation classes: successful plain trials vs. successful sucrose and quinine trials.
- **Taste** – Is the taste flavor encoded? Relevant separation classes: successful sucrose trials vs. successful quinine trials. (We take only the success because in the failure the mice didn't grab the food).
- **Specific flavor** – Are the different flavors distinguished in the encoding? Relevant separation classes: multiclass separation between the sucrose, quinine, and plain trials, multiclass separation between the sucrose, quinine, and plain successful trials.
- Are flavors also encoded in the failure trials? Relevant separation classes: multiclass separation between the sucrose, quinine, and plain failure trials.

For each component, for each time window, and for each question, we train a logistic regression model where the features are component values at time samples in the time window and the classified data points are the different trials. The labels are the question's separation classes. Classification was performed separately for each experimental session.

<u>Comparing classification methods</u>:

We tested two classification models - SVM and logistic regression. For each experimental session and for each question, we calculated the maximum accuracy of each of the methods. We finally chose logistic regression as it led to higher accuracies in most cases.

<u>Permutation testing and thresholding</u>:

We used permutation testing to determine whether a specific component and time window data is relevant to separating the labels. We permuted the labels 100 times and evaluated the mean and STD values of the accuracy. We then compared the actual accuracies to $\mu + 3\sigma$ where $\mu = \frac{1}{P}\sum_{p=1}^{P} ACC_p$ and σ is the standard deviation of the permutation accuracies $\sigma = \sqrt{\frac{1}{P}\sum_{p=1}^{P}\left(ACC_p - \mu\right)^2}$.

Based on the permutation testing results, we created an indicator tensor I_C of size $K \times W \times Q$ and I_N of size $N \times W \times Q$ indicating statistically significant component/neuron relevance to the questions in the specific time windows where the neuron's relevance is determined through its contribution to relevant components.

<u>Explained variance fraction</u>:

We calculated the explained variance of each component using the formula: $explained\ variance = 1 - \frac{\sum_{i=1}^{S\cdot T}\sum_{j=1}^{N}\left(R_{ij}-X_{ij}\right)^2}{\sum_{i=1}^{S\cdot T}\sum_{j=1}^{N}\left(X_{ij}\right)^2}$ where $X \in \mathbb{R}^{S\cdot T\times N}$ is the data matrix and $R \in \mathbb{R}^{S\cdot T\times N}$ is the reconstruction matrix: $R \in \mathbb{R}^{S\cdot T\times N}$.

Using the indicator tensor, I_C, we calculated for every question and time window, the fraction of the sum of explained variance of the components that encoded the answer to the question (See supplementary Table 1).

<u>Coding neurons fraction</u>:

The fraction of cells encoding a specific question in every time window is evaluated by averaging the first dimension of the indicator tensor I_N (See Figures 5,6, Supplementary Figure 1).

<u>Coding neurons emergence</u>:

Changes in encoding across experiment sessions are calculated by the Hamming distance between neurons populations that encode a given question over a specified time window across different experimental session types. Formally, for a time window w, and a question q

$$F_{ee'} = \frac{1}{\widetilde{N}}\sum_{n=1}^{N} I\{I_{N,e}[n,w,q] \neq I_{N,e'}[n,w,q]\} \ ; \ I\{I_{N,e}[n,w,q] \neq I_{N,e'}[n,w,q]\} = \begin{cases} 1 & I_{N,e}[n,w,q] \neq I_{N,e'}[n,w,q] \\ 0 & I_{N,e}[n,w,q] = I_{N,e'}[n,w,q] \end{cases}$$

Where $I_{N,e}[n,w,q]$ is n,w,q entry of session e's indicator tensor I_N, $F \in \mathbb{R}^{E\times E}$ is the emergence matrix and E is the number of experimental sessions (See full matrix in Supplementary Figures 2-4 and comparisons to the first training session in Figure 7).

2.2.3. Network Structure

<u>Degree centrality</u>:

We evaluated the correlation between cells per time segment. Given a correlation matrix, C, we computed the degree of each cell, n, as: $degree_n = \sum_{n'=1}^{N} C_{nn'}$. Changes in network structure across experiment sessions are calculated by the Euclidean distance between all pairs of degree vectors across experimental sessions yielding an $E \times E$ distance matrix. (See the full matrix in Supplementary Figure 5 and comparisons to the first training session in Figure 8).

<u>Correspondence between correlation degree and encoding</u>:

To examine the relationship between network structure (centrality scores) and encoding, we calculated Spearman's correlation rank between the neurons' degree and their encoding, quantified by relative accuracy, for every session, time window, and question (See Figure 9, Supplementary Figure 6).

Mathematically, we define the relative accuracy of a classifier as $ACC_{rel} = \frac{ACC-chance}{chance}$ where $chance = max\left\{\frac{labels==1}{labels}, \frac{labels==0}{labels}\right\}$, and a neuron's relative accuracy as a sum of relative accuracies of the relevant classifiers.

$$Spearman = 1 - \frac{6\Sigma d_i^2}{N(N^2-1)}\ ; d_i = R(A_i) - R(D_i)$$

Where $A \in \mathbb{R}^N$ is the relative accuracy vector, $D \in \mathbb{R}^N$ is the correlation degree vector and values A_i, D_i are converted to ranks $R(A_i), R(D_i)$.

3. Results and discussion

3.1. Fraction of encoding neurons

The fraction of encoding neurons indicates the proportion of the network participating in communicating specific information during a particular time window and experiment condition.

Fig. 5 presents the fraction of cells reporting binary outcomes (success vs. failure). Comparing the fraction of cells across the experimental conditions (marked with colors in Fig. 5) leads to a bell-shaped curve: in the early stages of the experiment ('train' and 'first' sessions), there is an increase in the encoding as the experiment progresses, indicating familiarity and understanding of the task. In later stages ('ongoing'), the network seems to engage in other processes beyond encoding the binary outcome (such as flavors), and in the 'ongoing random' sessions, encoding likely decreases due to the focus on flavor order and the flavor instability in the sequence.

We note that comparing the fraction of different questions can be biased. Analyzing a particular experimental session, the size of the classification data set is different for any question. For example, all the trials are relevant for comparing success vs. failure while only the sucrose trials are relevant for sucrose success vs. sucrose failure. Hence, we should be careful comparing the different questions related results and base the comparisons mainly on same question results. Most notably, the overall encoding of success and failure (Fig 5.A) is significantly higher than the others, likely because more data was available for this comparison, as it included all trials. This could explain the higher values since more classification results met the threshold, leading to more components being counted as encoded, which increased the fraction of encoding neurons.

We observe 15-20% encoding neurons in 'end' time window (Fig 5.A), compared to previous studies that showed 10% in analyses conducted per cell. This result is reasonable since our analysis considers the neuron population rather than single isolated neurons, and the neurons' interactions reveal additional encoding neurons. These results do not suffer from the above-mentioned bias, as the classification is based on all the trials.

Interestingly, we detect a small sub-population reporting outcome during the 'preparatory' time window, indicating some effect of the initial state of the network on performance in the upcoming trial. Once movement starts, the encoding of success vs. failure becomes more prominent where the highest values are achieved in the 'consumption' time window (2-5 seconds after the go cue).

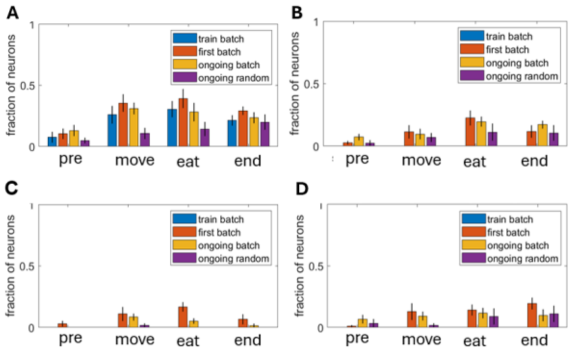


***Figure 5**: Fraction of neurons encoding: A. Success vs. failure. B. Success vs. failure for sucrose trials. C. Success vs. failure for quinine trials. D. Success vs. failure for grain and regular trials. For A-D the x-axis indicates the time segment within a trial and colors indicate session type.*

In Fig. 6 we present the reward-value encoding (flavors): novelty (grain vs. quinine and sucrose), taste (grain and sucrose vs. quinine) and specific flavor encoding. Importantly, the bias in Fig. 5 does not exist in these results because the classification tasks are based on all the flavors. For example, in the taste question we divide all the flavors into tasty and aversive ones hence all trials are included.

Observing the different experimental session types related results, the 'first' type (orange) shows no significant flavor classification, with a low number of neurons in the relevant columns. This result aligns with findings from other analyses [3] and [5]. In contrast, during 'ongoing' sessions (yellow and purple), flavor differences are clearly reported across all comparisons. Comparing 'first' to 'ongoing' session, we see relatively significant encoding for novelty in the 'first' sessions, unlike the taste. In the 'ongoing random' sessions compared to 'ongoing batch', in Fig. 5 we observe a decline in binary outcome encoding, but here, we see a rise in flavor encoding during the 'consumption' time window.

Novelty encoding is less significant compared to other classifications, with lower fraction of encoding neurons. Out of the three comparisons, the fraction of the network dedicated to encoding taste is the highest, suggesting that this distinction may be more meaningful than new taste exposure or differentiating flavors. Clear classification of specific flavors emerges only in later stages of the experiment (ongoing).

As expected, the strongest encoding occurs in the time windows after the animal tastes the flavors (2 seconds after the go cue and forward) and in the 'ongoing' trials where it is familiar with the flavors. In the 'preparatory' time window, there is no significant encoding for the novelty question (Fig 6.A), which is reasonable given the animal had not tasted it yet. However, significant encoding is observed in the 'preparatory' time window for taste and for specific flavor differences (Fig 6.B, C), especially for 'ongoing batch' sessions. In 'ongoing batch' understanding of the experimental set up and memory of previous trials allows the animal to predict the flavor before tasting quite well, which can explain this phenomenon.

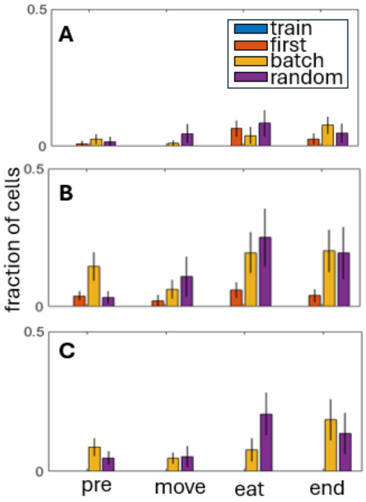


***Figure 6**: Fraction of neurons encoding: A. Novelty (grain vs. quinine and sucrose). B. Taste (grain and sucrose vs. quinine). C. Specific flavors. For A-C x axis indicates time segment within a trial and colors indicate session type.*

3.2. Explained variance fraction

The explained variance fraction measures the proportion of the total variance in the original dataset that is explained by the encoding components. In our current analysis, the explained variance is around 3% (Supplementary Table 1). This may seem little but in context, applying SPCA leaves us with 20-30%, so the result obtains 10% of that explained variance. Furthermore, in Fig. 5 we showed that more neurons are captured compared to prior studies.

3.3. Encoding neurons emergence

To examine the emergence of the encoding neurons as the experiments progress in time, we quantify, using Hamming distance, dissimilarity between populations that encode a given question over a specified time window across different experimental session types. The comparison is against the first 'train' experimental session.

In Fig. 7 we present results for the question: success vs. failure (binary outcome), examining two time windows – the 'preparatory' (red in the graph) and the 'consumption' (blue in the graph). The distances are calculated relative to the 'train'.

The 'preparatory' time window (red) shows non-significant results with a large standard deviation. The 'consumption' time window (blue) exhibits an increase in distance as the experiment progresses, with a significantly smaller standard deviation. The distance values for the 'preparatory' time window are lower than those for the 'consumption' time window. Since the 'preparatory' time window is before the tasting, it is expected to be more similar, hence the smaller distance.

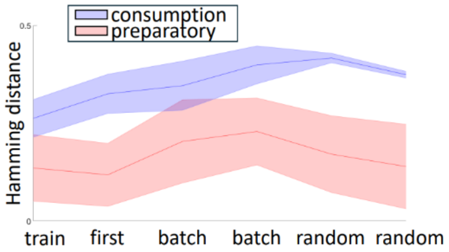


***Figure 7**: Hamming distance between the first sessions and the other sessions' neurons encoding success vs. failure in 'preparatory' [start,-1] (red) and 'consumption' [2-5] (blue) time window. The x-axis indicates the compared experimental session type.*

3.4. Correlation degree

To reveal trends in the network structure, we used Euclidean distance to quantify the dissimilarity between neurons' centrality in different experimental session types.

Fig. 8 presents the distance between neurons' correlation degrees in the 'train' experimental session and the other types. The calculation is performed separately for every time window (indicated by colors).

As the experiment advances, there is a notable change in the neuronal network activity, which increases the distance between correlation scores. The most significant variations are between 'train' and 'ongoing random' sessions, indicating differences in the nature of the experimental session, flavor, prior knowledge, differences between sessions, and the animal's prediction abilities.

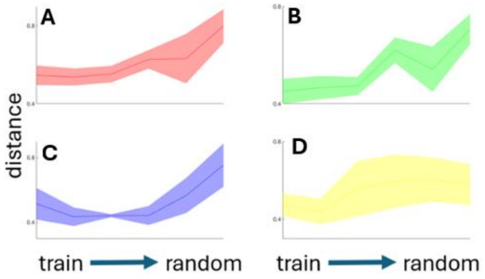


***Figure 8**: Euclidean distances between the first session's and the other sessions' neurons' correlation degrees in A. Preparatory B. Movement C. Consumption D. End of trial time window. For A-D x axis indicates the compared experimental session type.*

3.5. Spearman correlation rank

Here, we explore the relevance between the importance of sub-populations to encoding and their centrality in the overall network, through quantifying, using spearman correlation rank, the accordance between the neurons' centrality and encoding for every session, time window and question.

Fig. 9 presents Spearman correlation ranks between the correlation degrees and encoding populations for each time window and experimental session type, with white indicating non-significance or missing data. We focus on three questions: binary outcome, novelty and taste.

For the binary outcome (Fig 9.A), higher values are mainly observed in the early stages of the experiment before exposure to flavors, indicating that central neurons in the network encode binary outcome in these experiment stages. For novelty encoding (Fig 9. B), the first exposure to flavors shows more relevance, with subsequent parts of the experiment showing less significance (white squares). Lastly, the encoding of taste (Fig 9. C), displays a strong, persistent correspondence between connectivity and flavor encoding throughout most of the experiment.

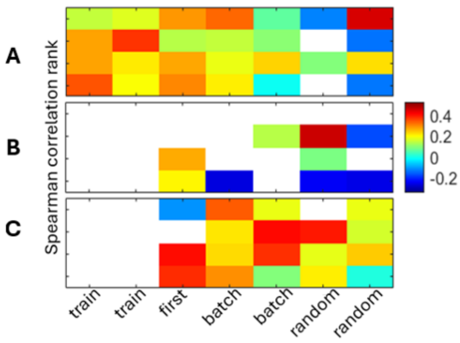


***Figure 9**: Spearman correlation ranks between the correlation degrees and encoding populations A. binary outcome (success vs. failure), B. novelty (grain vs. quinine and sucrose) C. taste (grain and sucrose vs. quinine). For A-C x axis indicates the experimental session type and y axis indicates the time window. White indicates non-significance.*

4. Conclusions

We see that the network changes with exposure to flavors (Fig. 8). The reporting of binary outcome by the neuronal network in layer 2-3 remains throughout all stages of the experiment, even after exposure to different flavors (Fig. 5). However, the way binary outcomes are reported changes upon flavor exposure (Fig. 5, Fig. 7). The encoding of flavors develops progressively. Initially, there is a distinction between encoding novelty, followed by taste, and ultimately encoding specific flavors (Fig. 6). Similarly, correspondence between network structure (centrality scores) and encoding changes progressively. Initially, binary outcome encoding matches the centrality, then novelty, and lastly taste (Fig. 9). Interestingly, it's noted that the network begins encoding flavors even before the initiation of movement, when the flavors are organized sequentially (Fig. 6).

5. Acknowledgements

Guidance and support from Dr. Hadas Benisty.

Imaging data was collected by Zohar Lotan from Jacky Shiller's - Technion Faculty of Medicine.

Guidance and support from Gaia Levin.

6. References

[1] Levy, S., Lavzin, M., Benisty, H., Ghanayim, A., Dubin, U., Achvat, S., ... & Schiller, J. (2020). Cell-type-specific outcome representation in the primary motor cortex. Neuron, 107(5), 954-971. Hantman,3, * and Jackie Schiller1,5

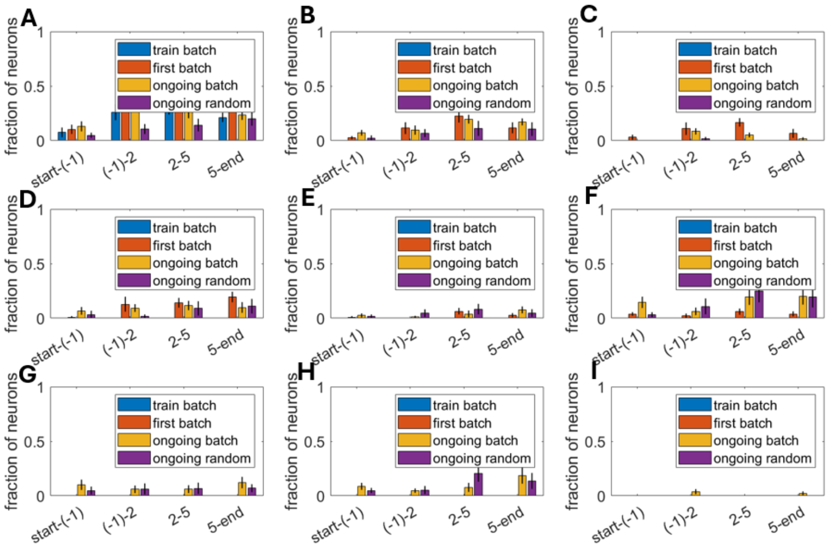
[2] Ghanayim, A., Benisty, H., Cohen-Rimon, A., Schwartz, S., Talmon, R., & Schiller, J. (2023). VTA projections to M1 are essential for reorganization of layer 2-3 network dynamics underlying motor learning. bioRxiv, 2023-11.

[3] Dynamic representation of task variables by layer 2-3 neurons in the primary motor cortex Research thesis, Zohar Lotan

[4] Lin, X. X., Nieder, A., & Jacob, S. N. (2023). The neuronal implementation of representational geometry in primate prefrontal cortex. Science Advances, 9(50), eadh8685.

[5] Sristi, R.D., Lindenbaum, O., Lifshitz, S., Lavzin, M., Schiller, J., Mishne, G. &amp; Benisty, H.. (2024). Contextual Feature Selection with Conditional Stochastic Gates. Proceedings of the 41st International Conference on Machine Learning, in Proceedings of Machine Learning Research 235:46375-46392

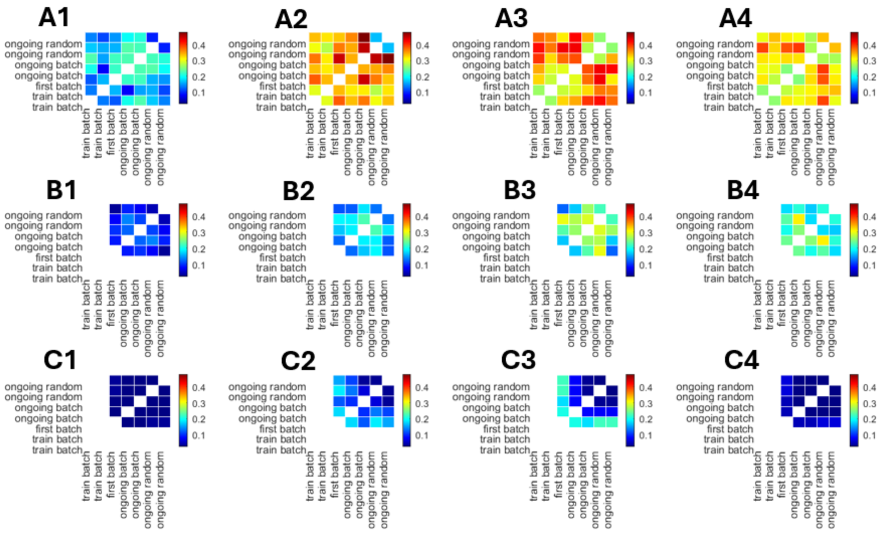
7. Appendix



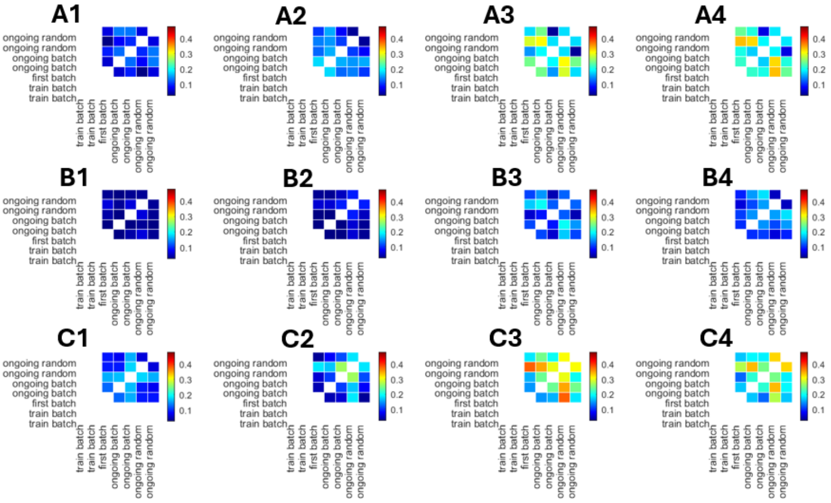
Supplementary Figure 1: *Fraction of neurons encoding: A. Success vs. failure. B. Success vs. failure for sucrose trials. C. Success vs. failure for quinine trials. D. Success vs. failure for grain and regular trials. E. Novelty (grain vs. quinine and sucrose). F. Taste (grain and sucrose vs. quinine). G. Specific flavors. H. Specific flavors success trials. I. Specific flavors failure trials. For A-I x axis indicates the time segment within a trial and colors indicate the session type.*

		Success vs. failure	Success vs. failure	Success vs. failure	Success vs. failure	Novelty (grain vs.	Taste (grain and	Specific flavors	Specific flavors	Specific flavors
train	preperatory	0.0126 +- 0.0076								
	motion	0.0418 +- 0.0135								
	consumption	0.0511 +- 0.0134								
	end	0.033 +- 0.0072								
first	preperatory	0.0136 +- 0.0056	0.003 +- 0.002	0.005 +- 0.004	0.0009 +- 0.0009	0.0007 +- 0.0007	0.004 +- 0.002	0 +- 0	0 +- 0	0 +- 0
	motion	0.0516 +- 0.0139	0.013 +- 0.006	0.012 +- 0.007	0.018 +- 0.01	0 +- 0	0.002 +- 0.001	0 +- 0	0 +- 0	0 +- 0
	consumption	0.0601 +- 0.0184	0.03 +- 0.008	0.019 +- 0.005	0.018 +- 0.007	0.008 +- 0.003	0.008 +- 0.003	0 +- 0	0 +- 0	0 +- 0
	end	0.0397 +- 0.0064	0.014 +- 0.007	0.01 +- 0.006	0.024 +- 0.008	0.003 +- 0.002	0.006 +- 0.003	0 +- 0	0 +- 0	0 +- 0
ongoing batch	preperatory	0.0158 +- 0.0053	0.008 +- 0.003	0 +- 0	0.007 +- 0.004	0.003 +- 0.002	0.02 +- 0.007	0.016 +- 0.007	0.013 +- 0.004	0 +- 0
	motion	0.041 +- 0.0064	0.011 +- 0.005	0.009 +- 0.003	0.01 +- 0.004	0.0008 +- 0.0009	0.008 +- 0.004	0.007 +- 0.004	0.005 +- 0.002	0.006 +- 0.004
	consumption	0.0411 +- 0.0104	0.029 +- 0.005	0.004 +- 0.002	0.014 +- 0.005	0.004 +- 0.003	0.03 +- 0.01	0.008 +- 0.005	0.012 +- 0.006	0 +- 0
	end	0.0321 +- 0.0061	0.023 +- 0.005	0.002 +- 0.002	0.014 +- 0.007	0.007 +- 0.002	0.03 +- 0.01	0.016 +- 0.006	0.03 +- 0.01	0.003 +- 0.002
ongoing random	preperatory	0.0056 +- 0.003	0.003 +- 0.003	0 +- 0	0.004 +- 0.004	0.001 +- 0.001	0.003 +- 0.002	0.005 +- 0.004	0.005 +- 0.003	0 +- 0
	motion	0.0122 +- 0.005	0.007 +- 0.004	0.001 +- 0.001	0.002 +- 0.002	0.006 +- 0.005	0.02 +- 0.01	0.008 +- 0.006	0.007 +- 0.005	0 +- 0
	consumption	0.0207 +- 0.0092	0.014 +- 0.008	0 +- 0	0.01 +- 0.01	0.012 +- 0.007	0.04 +- 0.02	0.011 +- 0.008	0.03 +- 0.01	0 +- 0
	end	0.0284 +- 0.0108	0.02 +- 0.01	0 +- 0	0.02 +- 0.01	0.007 +- 0.005	0.03 +- 0.02	0.009 +- 0.005	0.02 +- 0.01	0 +- 0

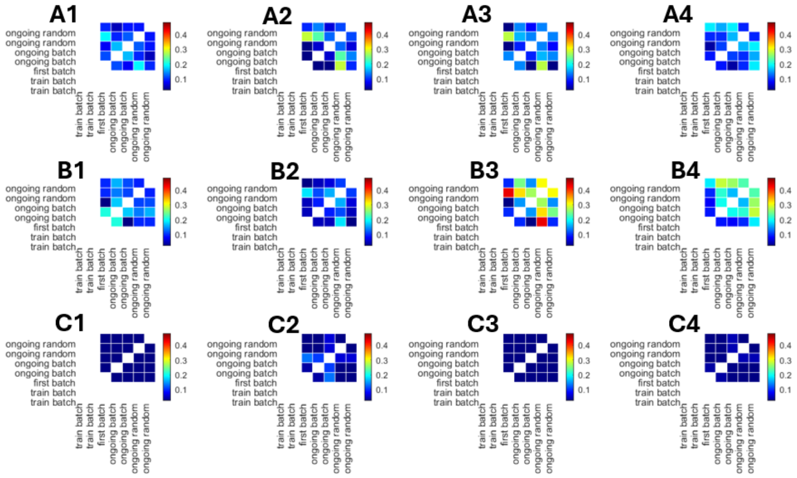
Supplementary Table 1: *fraction of explained variance for every question, time window and session type.*



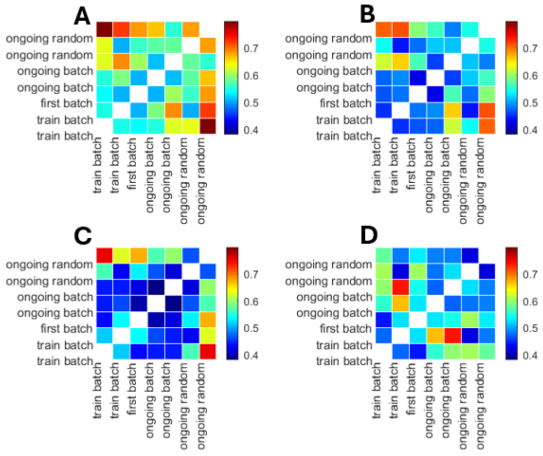
Supplementary Figure 2: *Hamming distance between the sessions' neurons encoding A. Success vs. failure. B. Success vs. failure for sucrose trials. C. Success vs. failure for quinine trials. in 1.'preparatory' 2. 'Movement'3.'consumption' 4. 'end' time window. The x and y axes indicate the compared experimental session types.*



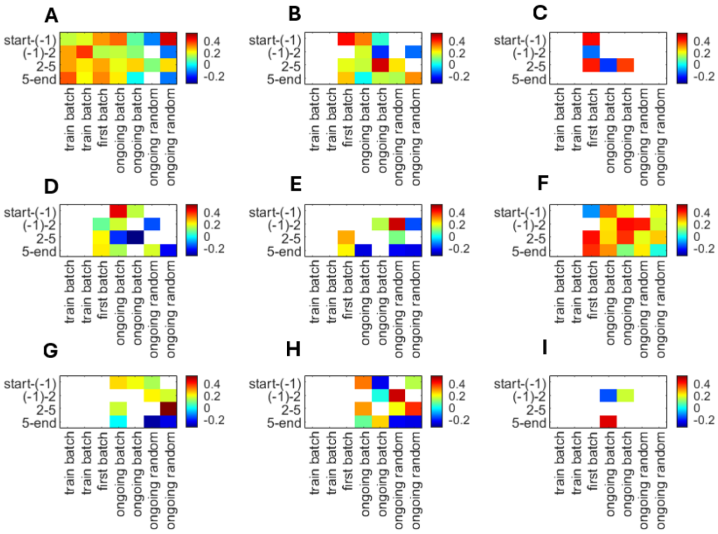
Supplementary Figure 3: Hamming distance between the sessions' neurons encoding A. Success vs. failure for grain and regular trials. B. Novelty (grain vs. quinine and sucrose). C. Taste (grain and sucrose vs. quinine). in 1.'preparatory' 2. 'Movement'3.'consumption' 4. 'end' time window. The x and y axes indicate the compared experimental session types.



Supplementary Figure 4: Hamming distance between the sessions' neurons encoding A. Specific flavors. B. Specific flavors success trials. C. Specific flavors failure trials. The x and y axes indicate the compared experimental session types.



***Supplementary Figure 5**: Euclidean distances between the sessions' neurons' correlation degrees in A. Preparatory B. Movement C. Consumption D. End of trial time window. For A-D the x and y axes indicate the compared experimental session types.*



***Supplementary Figure 6**: Spearman correlation ranks between the correlation degrees and encoding populations A. Success vs. failure. B. Success vs. failure for sucrose trials. C. Success vs. failure for quinine trials. D. Success vs. failure for grain and regular trials. E. Novelty (grain vs. quinine and sucrose). F. Taste (grain and sucrose vs. quinine). G. Specific flavors. H. Specific flavors success trials. I. Specific flavors failure trials. For A-I x axis indicates the experimental session type and y axis indicates the time window. White indicates non-significance.*