Novelty detection by density estimation in the fruit fly olfactory circuit

Summary– The capacity to recognize an unfamiliar situation is essential for survival. The fruit fly olfactory novelty detection circuit (Aso *et al.*, 2014) affords a tractable model for discerning computational principles of this ubiquitous ability, and, as this circuit has been extensively characterized, theorists have well-understood constraints to inform development of biologically-plausible algorithms. Stewart *et al.* (2023) hypothesize that in insects, synapses between mushroom body Kenyon Cells and the MBON $\alpha'3$ neuron assess probability of an input given some history, by means of an intermediate projection to a high-dimensional vector space. However, it is unclear whether this algorithm would work when subject to the extreme constraints of the brain of the fruit fly specifically, one of the most widely-used animal models in neuroscience. To address this uncertainty, here we characterize a variant of the model with size, sparsity, and connectivity corresponding to the *Drosophila melanogaster* circuit. We observe that incorporating these biological details reproduces aspects of Kenyon Cell response patterns as seen in actual recordings. Moreover, there is a monotonic relationship between probability and network output, which would support its application as a novelty detector. Characterizing model performance on a novelty detection task as a function of network size and sparsity, we find that the fruit fly circuit model acts as an ideal classifier, as assessed with the area under the curve of the receiver operator characteristic. These findings simultaneously uphold the interpretation of the fruit fly mushroom body as performing novelty detection via probability estimation and invites the application of the approach to other species, towards a deeper understanding of novelty detection across the Arthropod phylum.

Additional Detail– Stewart *et al.* model novelty detection by the insect mushroom body circuit. Input data (of any dimensionality) is projected onto a manifold lying on the surface of a unit hypersphere with fixed dimensionality $D = 1024$. Here, a feature of the input space is assigned a randomly selected vector whose Fourier components have unit magnitude, which supports encoding of continuous-valued data (scaled by some factor λ) following the method of fractional binding (Komer *et al.*, 2019) by circular convolution. Multidimensional data is encoded by binding the results of each such mapping together (e.g. $< x, y > \rightarrow X^{x/\lambda} \circledast Y^{y/\lambda}$, where x, y is a point in a 2D input space, X, Y are the axis vectors selected for each feature, and $\circledast$ denotes the circular convolution operator). Samples in this latent representation are provided as input to a population of 50000 neurons (representing Kenyon cells, (KCs)), with preferred stimuli, or *encoders*, randomly sampled from the surface of the unit hypersphere. During the training phase, connection weights w to the MBON$\alpha'3$ neuron are updated using a simple learning rule based only on the presynaptic activity, a_i: $\frac{dw_j}{dt} = \frac{a_i}{\sum_j a_j}$. The probability of a sample given the dataset is then estimated as a weighted sum of the activities in the KC layer: $y = A \cdot w^T$, where A is a $k \times n$ matrix with k the number of samples and n the number of neurons in

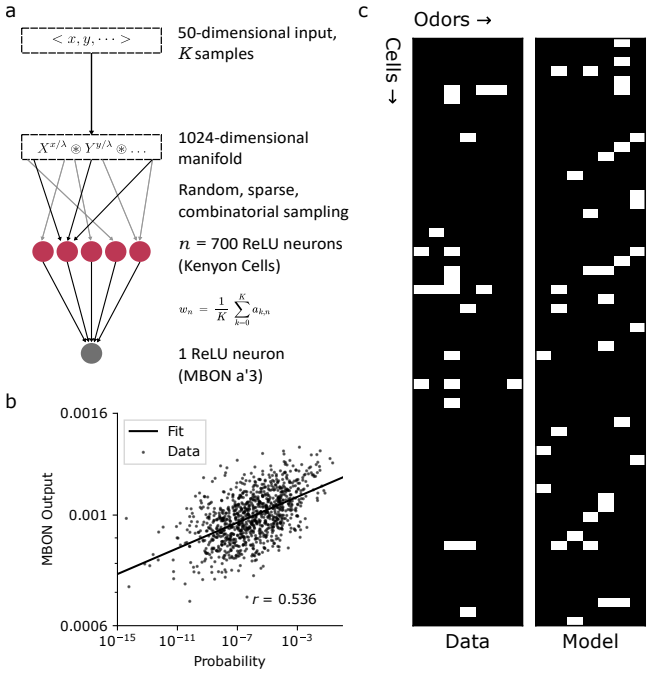

Fig. 1: **Simulated mushroom body activity reproduces features of Kenyon Cell and MBON $\alpha'3$ response patterns. a.** Network model. **b.** Scatterplot revealing a relationship between the logarithms of network output and ground-truth probability of sample inputs. **c.** Comparison of activity from imaged and simulated KCs. In the data, white bars indicate that a cell's response is statistically-significant, on average, across odor presentations. In the model, white bars indicate that the cell is active. 62 cells were randomly selected from both the data and simulations for plotting. Data reproduced from Srinivasan *et al.* (2023).

the population, and $\boldsymbol{w^T}$ is a $n \times 1$ column vector. In other words, the output of the network will be large for familiar samples, and small for unfamiliar or novel samples.

The significance of Stewart's model lies in the assumption that KCs act on a high-dimensional vector space, enabling a direct probabilistic interpretation: In this latent space, the dot product between two vectors approximates a normalized sinc function, which, after rectification, induces an admissible kernel for density estimation (Furlong & Eliasmith, 2022). As the response of a KC is the rectified result of the dot product between its preferred direction and the input plus some bias, the weighted sum of population activity of such neurons can be interpreted as performing density estimation.

Here we characterize the model's performance when constrained by the neurobiology of fruit fly olfaction. The model is depicted in Fig. 1a. Following the convention from Dasgupta *et al*, we take olfactory input to be samples from a 50-dimensional space, corresponding to the 50 different types of olfactory receptor neurons (Dasgupta *et al.*, 2018). We set the KC population size to 700, corresponding to 350 neurons converging onto one MBONα'3 neuron per hemisphere (Dasgupta *et al.*, 2018), both of which contribute to behavior. We impose a strong negative bias, ξ, on rectified linear neurons to mimic KCs' high thresholds and enforce a sparsity of ρ = 0.06 (KC response probability is 6% on average across stimuli and cells; Turner *et al.*, 2008). We also incorporate details on the connectivity patterns between projection neurons (PNs) and KCs. Specifically, KCs receive input from 1-10 PNs (Gruntman & Turner, 2013). We model this by first sampling vectors from the manifold of the latent representation, and constructing KC encoders by equally-weighted combinations of these vectors. The number of components in a vector "bundle" for a given cell, in turn, were drawn from a discrete distribution over the range [1,7] with probabilities [0.3, 0.25, 0.2, 0.1, 0.05, 0.05, 0.05], in rough agreement with the actual data (Gruntman & Turner, 2013). For simplicity, we evaluate a time-flattened version of the model with the weight for the n^{th} neuron: $w_n = \frac{1}{K}\sum_{k=0}^{K} a_{k,n}$, and use neurons with rectified linear tuning curves. Network output exhibits a monotonically increasing relationship with probability (Fig. 1b). The simulated KC population exhibits similar overall sparseness, and includes cells that respond to one or more odors as seen in the real data (Fig. 1c).

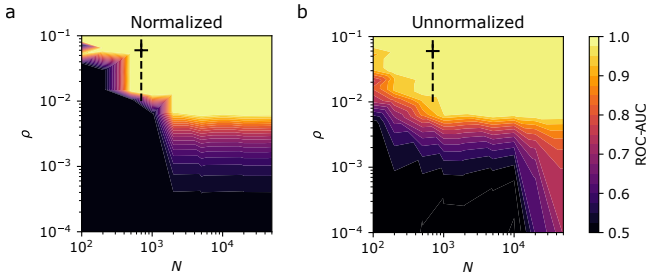

Fig. 2: **Incorporating neurobiological constraints maintains strong performance of the novelty detection network.** Performance landscape as measured by the area under the curve of the receiver-operator characteristic (AUC-ROC) for the novelty detection network over the sparsity (ρ) and KC population size (N) parameter space with **a.** Activity normalized across KCs and **b.** Activity left unnormalized. The $+$ marker indicates the actual average operating point, and the dotted line indicates a range of sparsity levels observed.

We next characterized model performance on a novelty detection task (Fig. 2) in which 1000 samples were provided to the network, 85% of which represented familiar stimuli (drawn from a 50-dimensional unit-variance, origin-centered normal distribution) and 15% were novel or unfamiliar (drawn from a uniform distribution over a large range). Models using sparsities and network sizes observed in the fruit fly results in strong performance, demonstrating robustness of the model to the extreme biological constraints of this species (Fig. 2a). The novelty detection network performed strongly even when activities were left unnormalized (Fig. 2b), demonstrating that normalization is not strictly required by the model.

This work upholds the interpretation of the insect novelty detection circuit as a density estimator, and extends this interpretation to the fruit fly circuit. This work bears some resemblance to the model of Dasgupta et al, which interprets the novelty detection circuit as a kind of Bloom filter: the output of a traditional Bloom filter is binary, so the model is modified to account for the continuous-valued output of the MBONα'3 neuron. In contrast, the density estimation model we evaluate here *predicts* these characteristics.

Future work should evaluate model performance that incorporates the stochastic and dynamic spiking of KCs observed across Arthropods.

References

Aso, Y., Hattori, D., Yu, Y., Johnston, R. M., Iyer, N. A., Ngo, T.-T., Dionne, H., Abbott, L., Axel, R., Tanimoto, H., *et al.* (2014). *elife,* **3**, e04577.

Dasgupta, S., Sheehan, T. C., Stevens, C. F., & Navlakha, S. (2018). *Proceedings of the National Academy of Sciences,* **115** (51), 13093–13098.

Furlong, M. & Eliasmith, C. (2022). In: *Proceedings of the Annual Meeting of the Cognitive Science Society* volume 44 ,.

Gruntman, E. & Turner, G. C. (2013). *Nature neuroscience,* **16** (12), 1821–1829.

Komer, B., Stewart, T. C., Voelker, A. R., & Eliasmith, C. (2019). In: *41st Annual Meeting of the Cognitive Science Society* , Montreal, QC: Cognitive Science Society.

Srinivasan, S., Daste, S., Modi, M. N., Turner, G. C., Fleischmann, A., & Navlakha, S. (2023). *PLoS Biology,* **21** (10), e3002206.

Stewart, T., Furlong, P., Simone, K., Bartlett, M., & Orchard, J. (2023). In: *MathPsych/ICCM/EMPG 2023* ,.

Turner, G. C., Bazhenov, M., & Laurent, G. (2008). *Journal of neurophysiology,* **99** (2), 734–746.