The Emergence of 2-gram Neurons and Feature Binding in a Spiking Neural Network Model of the Human Visual Cortex

Supervisor: Greg Davis

Summary

The feature binding problem in vision concerns how the human visual system assembles discrete features into complete object representations that one consciously perceives, a question central to understanding the human brain. In this computational study, I tested the concept of 2-gram neurons, or neurons that encode pairwise feature conjunctions, as a novel solution to this problem. I constructed a spiking neural network model of the human visual cortex to serve as the in silico platform for this study. By presenting the cortex with intrinsically structured, object-like visual stimuli and updating it via unsupervised learning, I showed that neurons with 2-gram-like tuning properties emerged from the cortex. By probing the cortex at various retinal positions, I demonstrated that the emerged 2-gram neurons are sensitive to objects' spatial locations. By further decoding their firing responses, I revealed that the emerged 2-gram neurons cannot explicitly encode feature combinations in visual stimuli. Together, this study demonstrates the spontaneous emergence of 2-gram neurons in a neurobiologically realistic model of the human visual system and elucidates their potential and limitation as a novel mechanism for addressing the feature binding problem.

Table of Contents

Abbreviations

FBP: Feature binding problem

RDM: Representational dissimilarity matrix

1. Introduction

The feature binding problem in vision (FBP) concerns how the human visual system integrates separately encoded visual features, such as shape, colour, motion, etc., into holistic object representations that one consciously perceives (von der Malsburg, 1981). This problem arises from the experimental observations that neurons in the visual cortex are selectively tuned to distinct features at different retinal locations (Livingstone & Hubel, 1987; Nassi & Callaway, 2009; Taylor & Xu, 2022). Consequently, when multiple objects appear simultaneously in a visual scene, features from different objects would concurrently activate a pool of neurons across the cortex, creating a challenge for the visual system in differentiating which feature belongs to what object. This neural ambiguity, however, does not prevent human vision from accurately recognising objects and their feature compositions. Thus, it is assumed that a dedicated neural mechanism must exists to resolve this challenge (see Di Lollo, 2012 for opposing claims).

Fig 1.1 provides a straightforward illustration of the FBP. Consider a simplified human visual cortex that contains six neurons: two selectively tuned to the colours red and green, two to the shapes square and triangle, and two to the locations top and bottom, which replicates the visual system's feature selectivity. While this cortex is capable of encoding single objects (Fig 1.1a and Fig 1.1b), a binding problem arises when two objects are simultaneously presented as the visual input (Fig 1.1c), as the cortex can no longer unambiguously represent the combination of shape, colour, and location for each object in the scene. Note that, for simplicity, the presented objects in this illustration contain spatially overlapping features that can be bound trivially via locations (Treisman, 1988). However, multipart objects with hierarchical feature compositions, as those in naturalistic scenes, cannot be bound simply by spatial cues.

Understanding the generic binding mechanism that these complex objects require remains one of the biggest open questions about the human brain (Feldman, 2013).

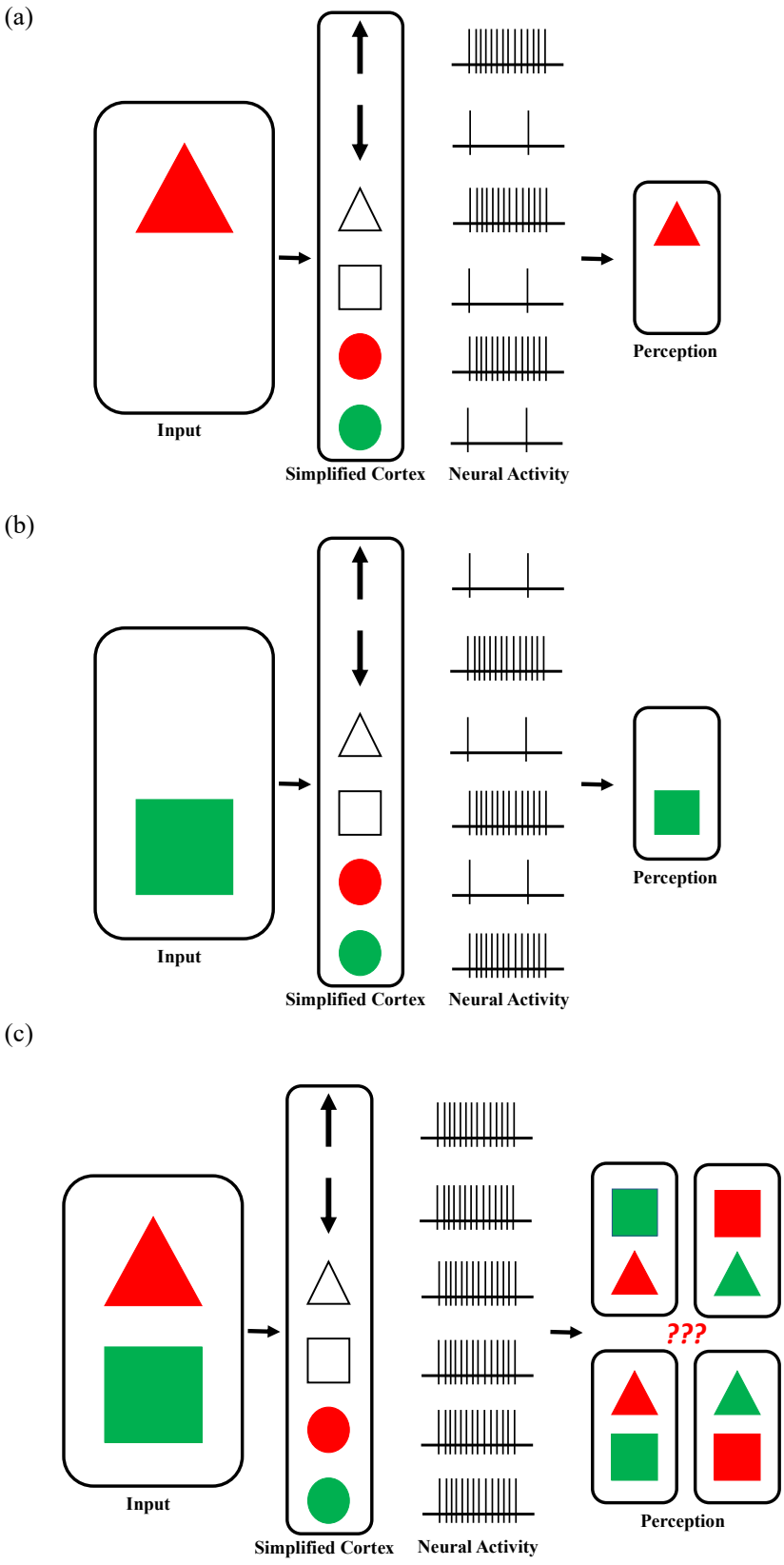

Fig 1.1: A classical illustration of the FBP by Rosenblatt (1962). A cortex with selectively tuned neurons can uniquely encode (a) a single red triangle at the top location, or (b) a single green square at the bottom location, but not (c) the simultaneous presentation of both objects. Icons in the cortex represent neurons with specific feature selectivity, with raster plots conceptually indicate levels of neural activation. Adapted from Velik (2012).

Since the popularisation of the FBP, several binding mechanisms have been proposed, yet each remains flawed or incomplete. The most prominent theories among them include *binding by hardcoded conjunctions* (Barlow, 1972; Riesenhuber & Poggio, 1999; Fig 1.2a), which uses neurons tuned to complete sets of feature conjunctions to directly represent feature combinations, *binding by synchrony* (von der Malsburg, 1981; Fig 1.2b), which applies synchronised neural oscillations to temporally group neurons encoding the same object, and *binding by rate-enhancement* (Treisman & Gelade, 1980; Reynolds & Desimone, 1999; Fig 1.2c), which utilises attention-induced neural activations to bind neurons activated by the same object. Despite some theoretical and experimental supports (hardcoded: Riesenhuber & Poggio, 1999; Mel, 1997; synchrony: Singer & Gray, 1995; Fries et al., 2007; rate-enhancement: Roelfsema et al., 1998; Jeurissen et al., 2016), all three solutions have faced criticisms regarding their neurobiological plausibility and the lack of clearly defined neural mechanisms (Roelfsema, 2023).

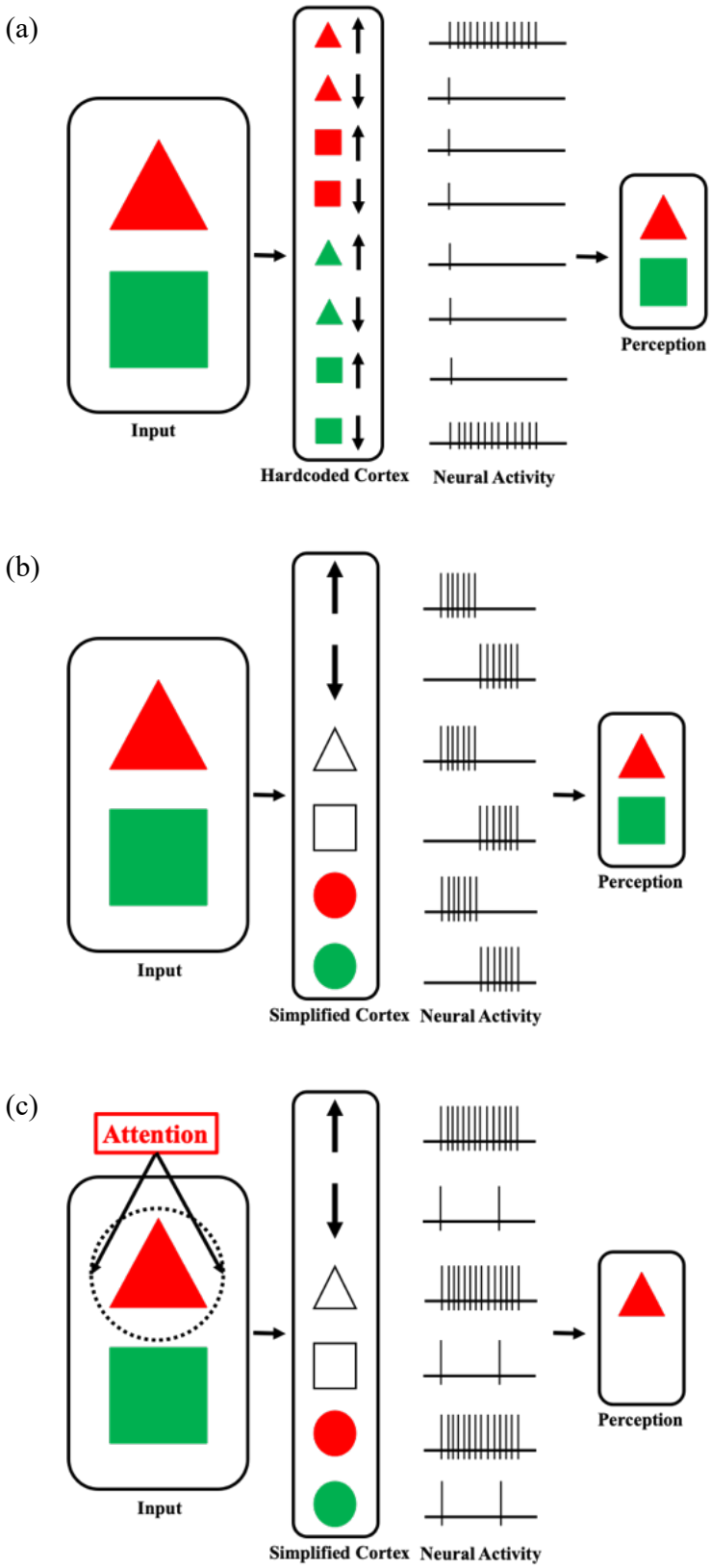

Fig 1.2: Solving the FBP via (a) binding by hardcoded conjunction, (b) binding by synchrony, and (c) binding by rate-enhancement. Icons in the cortex represent neurons with specific feature selectivity, with raster plots conceptually indicate the level of neural activations. Adapted from Velik (2012).

2-grams are pairs of adjacent elements extracted from a continuous sequence, originally introduced in linguistics for text analysis (Cavnar & Trenkle, 1994). By decomposing a word (e.g., *biology*) into a set of interconnected letter pairs (e.g., {*bi, io, ol, lo, og, gy*}), 2-grams allow the unambiguous and efficient representation of letter orders in common words (Mel & Fiser, 2000). Given the intricate similarity between assigning letters to words and binding features to objects, 2-grams could provide an alternative solution to the FBP (Fig 1.3). Rather than relying on hardcoded conjunctions or external binding mechanisms, the visual cortex could use 2-gram neurons to encode pairwise feature combinations in visual inputs, which, collectively, enable the unambiguous representation of each object's feature composition in the scene.

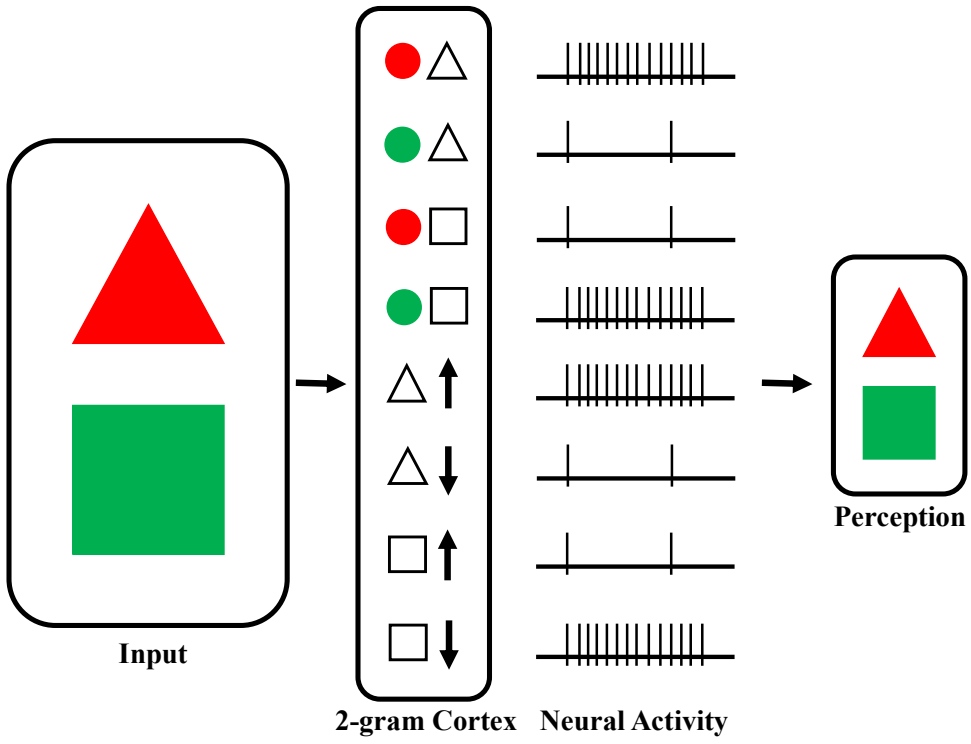

Fig 1.3: Solving the FBP by 2-gram neurons. Icons in the cortex represent neurons with specific feature selectivity, with raster plots conceptually indicate levels of neural activation.

2-grams are not entirely new to visual neuroscientists, as recent studies have explored 2-gram neurons, or their equivalents, as potential mechanisms for feature binding. Morita et al. (2010) demonstrated that rapid flashes of feature pairs could reliably induce illusions of feature conjunctions in both human observers and a connectionist network of 2-gram units. Furutate et al. (2019) showed that, using a perceptual conditioning task, 2-gram-like representations could accurately explain human performances on learning feature conjunctions. Schneegans & Bays (2017) further revealed that human recall errors in visual working memory are best predicted by a neural population encoding feature pairs. These studies have shown strong evidence for the involvement of 2-gram neurons in feature binding. However, several critical questions remain.

First, it is unclear how 2-gram neurons are formed in the visual system. Do they spontaneously emerge through the brain's intrinsic plasticity, similar to statistical learning in vision (Fiser & Lengyel, 2022), or are they hardwired through certain innate mechanisms? Second, it is unknown whether 2-gram neurons exhibit positional invariance, that is, whether they response invariantly to an object's location in the visual field, which is an important characteristic of human object recognition (DiCarlo & Cox, 2007). Third, while prior works have shown that 2-gram neurons can bind simple features like shape and colour, the ability of 2-gram neurons to encode more abstract feature combinations in real-world objects is still debatable. These questions were the main focus of the present study.

Beyond traditional experimental approaches, simulation studies have become an increasingly popular approach to investigating the human brain (Fan & Markram, 2019). By avoiding technical challenges and ethical concerns associated with in vivo neurophysiological experiments, computational models enable systems neuroscientists to pursue broader research

questions with more flexible experimental designs. Spiking neural networks, in particular, provide an excellent framework for in silico brain experiments (Maass, 1997). By incorporating biophysically realistic neurons that represent information via discrete spikes (Hodgkin & Huxley, 1952; Izhikevich, 2003; Burkitt, 2006), spiking neural networks replicate key aspects of the human brain (Yamazaki et al., 2022) and have been applied successfully in recent studies on the FBP (Miconi & VanRullen, 2010; Martin & von der Heydt, 2015; Isbister et al., 2018). This computational framework served as the main methodology for the present study.

In this study, I investigated the formation of 2-gram neurons in a spiking neural network model of the human visual cortex and examined their potentials in solving the FBP. In particular, I aimed to address the following questions:

1. Can 2-gram neurons spontaneously emerge in the in silico cortex when exposed to intrinsically structured, object-like visual stimuli?
2. Are these emerged 2-gram neurons, if any, invariant to stimulus locations in the visual field?
3. Do these emerged 2-gram neurons, if any, directly encode the abstract feature combinations of complex objects?

By presenting the in silico cortex with intrinsically structured, object-like visual stimuli and updating it via unsupervised learning, I showed that neurons with 2-gram-like properties emerge in the cortex. By probing the cortex at various retinal locations, I demonstrated that these neurons are not positionally invariant. Furthermore, by decoding the recorded neural responses, I revealed that these neurons do not explicitly encode feature combinations in visual stimuli. Together, this project demonstrates the spontaneous formation of 2-gram neurons in a

biologically realistic model of the human visual system and evaluated their potential and limitation as a novel mechanism for addressing the FBP.

2. Method

2.1 In Silico Visual Cortex

Inspired by Eguchi et al. (2018), I constructed a spiking neural network model of the human visual cortex that implemented hierarchical network architectures (Serre, 2012), leaky integrate-and-fire neurons (Burkitt, 2006), and spike-timing dependent plasticity (Perrinet et al., 2001) to simulate the structure, dynamics, and flexibility of a human visual system.

Network architecture. The cortex consisted of five two-dimensional cortical layers arranged in a bottom-up hierarchy, which approximated the human ventral visual pathway for object recognition (Kravitz et al., 2013; Fig 2.1). Layer 0 contained hardwired excitatory cells that preprocessed raw visual images. Layer 1 to 4 contained subpopulations of excitatory and inhibitory neurons that progressively propagated and transformed the preprocessed visual stimuli. In total, the cortex consisted of 4096 simple cells (4 for 32×32 retinal locations), 4096 excitatory neurons (32×32 for 4 layers), and 1024 inhibitory neurons (16×16 for 4 layers).

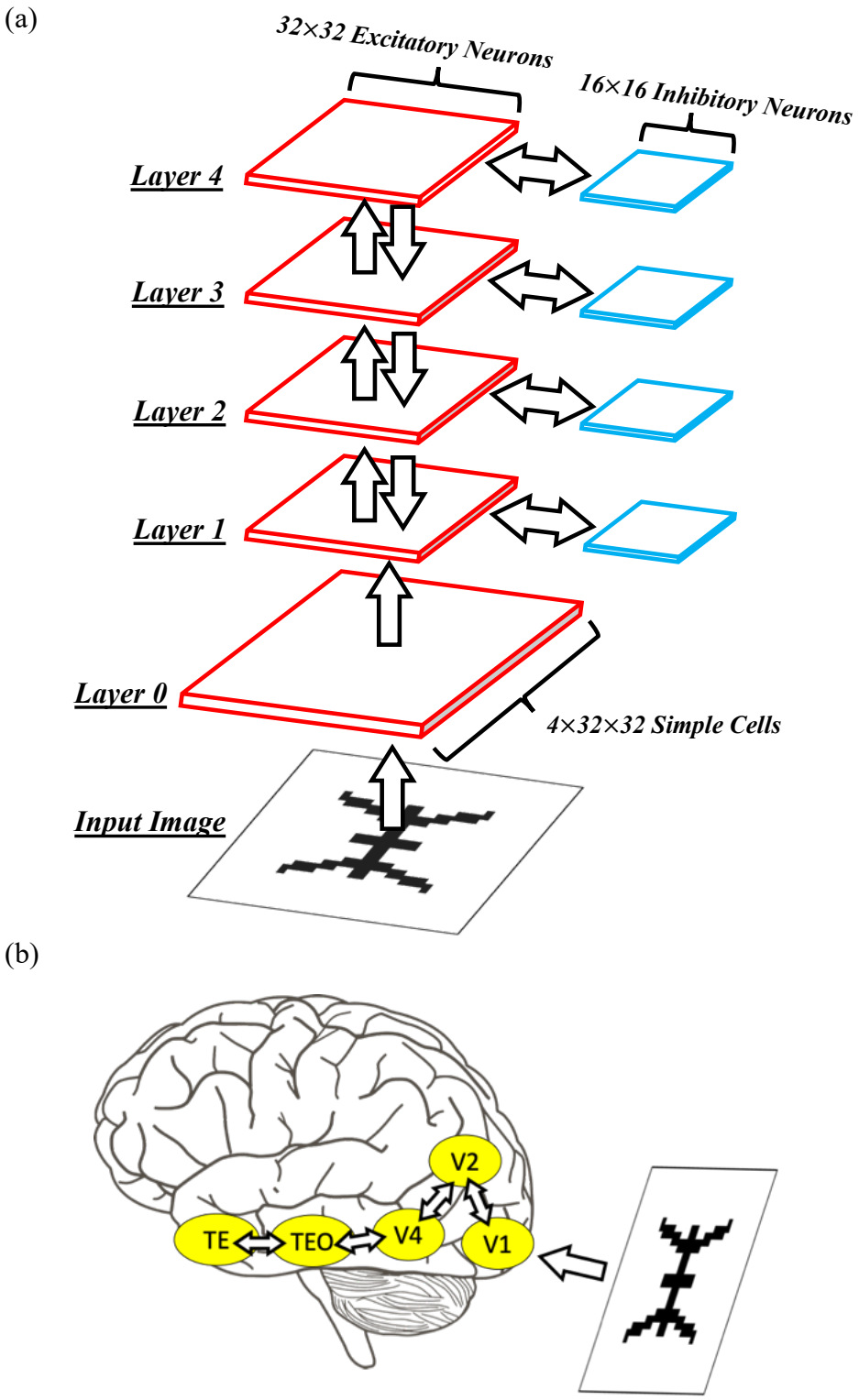

Fig 2.1: (a) Overall architecture of the in silico cortex and (b) its similarity with the human ventral visual pathway. Red and blue blocks indicate excitatory and inhibitory neural subpopulation, respectively. Arrows indicate permitted directions of information flow among neural subpopulations in the cortex. Layer 0 to 5 in the cortex respectively corresponded to V1, V2, V4, posterior inferior temporal cortex (TEO), and anterior inferior temporal cortex (TE) in the human visual system.

Synaptic connection. Neurons in the in silico cortex communicated through conductance-based synapses. Specifically, excitatory neurons formed feedforward and feedback connections with excitatory neurons in successive layers, as well as lateral connections with both excitatory and inhibitory neurons in the same layer. Inhibitory neurons formed only local lateral synaptic connections with excitatory neurons in the same layer. These connections were spatially confined, where neurons formed synapses primarily within a defined projection radius (Fig 2.2). Each presynaptic neuron selected its postsynaptic targets by sampling from a bivariate Gaussian distribution, with its presynaptic spatial coordinates as the mean and its projection radius as the covariance. The projection radius was gradually increased across cortical layers to replicate the expanding neural receptive fields observed along the ventral visual pathway (Kravitz et al., 2013).

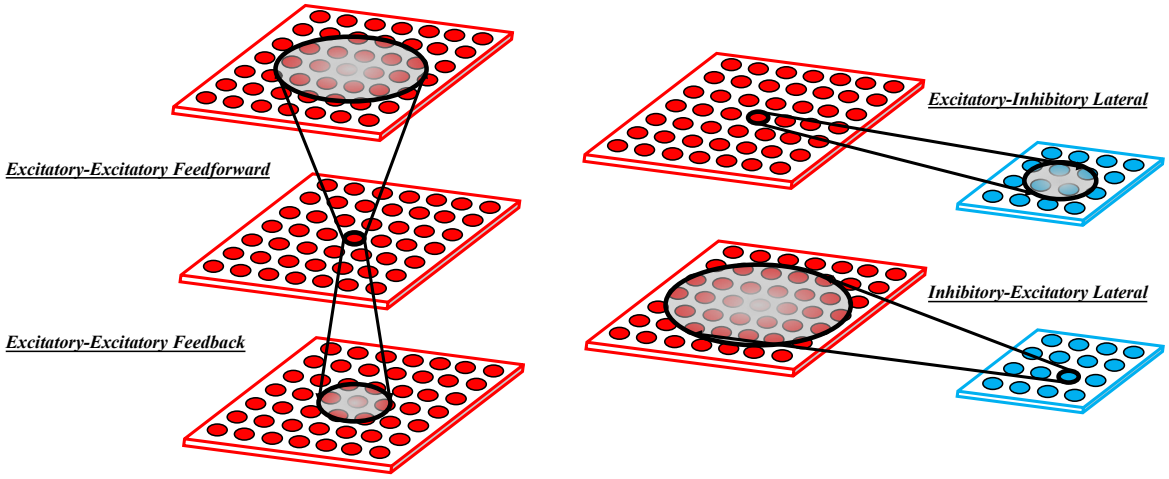

Fig 2.2: Synaptic connections among neurons in the in silico visual cortex. Red and blue circles indicate excitatory and inhibitory neurons, respectively, with presynaptic cells mark with dark edges and their postsynaptic projection radii mark with gray shaded areas.

Simple Cells. Cells in Layer 0 were modelled as excitatory simple cells (Hubel & Wiesel, 1962), which fire strongly to oriented edges in visual stimuli at specific retinal locations. When a light impulse occurred at a retinal coordinate (x, y), the firing rate of a simple cell, f, followed a modified Gabor filter (Kruizinga & Petkov, 1999):

$$f = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\bar{\lambda}} + \psi\right) \tag{1}$$

with

$$\begin{cases} x' = x \cdot \cos\theta + y \cdot \sin\theta \\ y' = y \cdot \cos\theta - x \cdot \sin\theta \\ \sigma = \sqrt{\frac{\ln 2}{2}} \cdot \frac{\bar{\lambda}(2^b + 1)}{\pi(2^b - 1)} \end{cases} \tag{2}$$

where θ is the cell's preferred edge orientation, σ and γ set the geometric properties of the cell's receptive field, $\bar{\lambda}$ and b control the cell's spatial frequency, and ψ determines the phase in the cell's response. For each of the 32×32 retinal locations, a group of simple cells with four different preferred orientation was deployed ($\theta = 0°$, 45°, 90°, and 135°, respectively).

Simple cells emitted all-or-none spike trains with equally spaced spikes as output, with interspike interval, T, determined by:

$$T = \frac{f_{\max}}{f}\omega \tag{3}$$

where $f_{\max}$ is the maximum f among the entire simple cell population, and ω represents the upper bound of a simple cell's firing intensity.

Neuron Equations. Excitatory and inhibitory neurons in layer 1 to 4 were modelled as leaky integrate-and-fire neurons, which replicated the biophysical properties of pyramidal cells in

mammalian neocortex (Jolivet et al., 2006). At time t, the membrane potential of a neuron i, denoted as $V_i(t)$, was updated via:

$$\tau_m^{\gamma} \frac{dV_i(t)}{dt} = V_0^{\gamma} - V_i(t) + R^{\gamma} I_i(t) \tag{4}$$

where τ_m^{γ} represents the membrane time constant, V_0^{γ} is the resting potential of the neuron, $I_i(t)$ is the total synaptic current received by neuron i at time t, R^{γ} is the membrane resistance, and γ index indicates the neuron type. This formulation ensured that $V_i(t)$ passively decayed toward the resting state V_0^{γ} while being actively driven by synaptic inputs.

Neurons emitted all-or-none spikes. At time t, the binary spike state of neuron i, $\delta_i(t)$, was computed via:

$$\delta_i(t) = \begin{cases} 1, & \text{if } V_i(t) > \Theta^{\gamma} \text{ and } \Delta\delta_i(t) > \tau_R \\ 0, & \text{elsewise} \end{cases} \tag{5}$$

where Θ^{γ} is the threshold potential, and $\Delta\delta_i$ is the time elapsed since the neuron's last spiking event. After a spike was emitted, the neuron entered an absolute refractory period that lasted for a duration of τ_R, during which $V_i(t)$ was fixed at the neuron's hyperpolarised potential, V_H^{γ}, and no spiking was allowed.

The current term $I_i(t)$ in Equation (4) was defined as the sum of all excitatory and inhibitory synaptic currents injected into the neuron i:

$$I_i(t) = \sum_j \left(\hat{V}^{\gamma} - V_i(t)\right) g_{ij}(t) \tag{6}$$

where $\hat{V}^{\gamma}$ is the synaptic reversal potential, and $g_{ij}(t)$ is the synaptic conductance from presynaptic neuron j to postsynaptic neuron i, which was updated via:

$$\frac{dg_{ij}(t)}{dt} = -\frac{g_{ij}(t)}{\tau_r} + \lambda \cdot \Delta g_{ij}(t)\delta_j(t) \tag{7}$$

where τ_r is the synaptic time constant, λ is the biological scaling factor, $\Delta g_{ij}(t)$ is the synaptic efficacy, and $\delta_j(t)$ represents the presynaptic spiking event as defined in Equation (5).

Unsupervised learning. Spike-timing dependent plasticity was applied to modify excitatory-excitatory synaptic efficacies based on the temporal correlation between presynaptic and postsynaptic spiking. Following Perrinet et al. (2001), this rule was implemented using the updating equation:

$$\frac{d\Delta g_{ij}(t)}{dt} = \rho\left[\left(1 - \Delta g_{ij}(t)\right) \cdot C_{ij}(t)\delta_i(t) - \Delta g_{ij}(t) \cdot D_i(t)\delta_j(t)\right] \tag{8}$$

where ρ is the learning rate, $C_{ij}(t)$ quantifies presynaptic activity, and $D_i(t)$ quantifies postsynaptic activity. The value of $\Delta g_{ij}(t)$ was clamped between [0, 1] to ensure numerical stability.

$C_{ij}(t)$ and $D_i(t)$ in Equation (8) were updated by:

$$\frac{dC_{ij}(t)}{dt} = -\frac{C_{ij}(t)}{\tau_C} + \alpha_C \left(1 - C_{ij}(t)\right) \cdot \delta_j(t) \tag{9}$$

and

$$\frac{dD_i(t)}{dt} = -\frac{D_i(t)}{\tau_D} + \alpha_D \left(1 - D_i(t)\right) \cdot \delta_i(t) \tag{10}$$

where τ_C and τ_D are the presynaptic and postsynaptic time constants, and α_C and α_D are proportionality constants that describe the presynaptic and postsynaptic learning resources. Conceptually, $C_{ij}(t)$ represented the concentration of glutamate neurotransmitter residue in the synaptic cleft, and $D_i(t)$ represented the number of unblocked glutamate receptors on the postsynaptic membrane. Both variables decayed passively and were activated asymptotically by presynaptic or postsynaptic spiking.

Equation (8) to (10) ensured that changes in synaptic efficacy were governed by spike timing. When a presynaptic neuron fired shortly before a postsynaptic neuron, the presynaptic term involving $C_{ij}(t)$ in Equation (8) dominated, which increased $\Delta g_{ij}(t)$. Conversely, when a postsynaptic neuron fired before a presynaptic neuron, the postsynaptic term involving $D_i(t)$ dominated, which decreased $\Delta g_{ij}(t)$. The magnitude of change in $\Delta g_{ij}(t)$ was inversely proportional to the interval between presynaptic and postsynaptic spikes. Hence, a shorter spiking interval, indicating a stronger neuronal correlation, produced a greater learning effect.

All neuronal and synaptic parameters were summarised in Appendix I, with values adopted from available neurophysiological measurements of the mammalian visual cortex (Perrinet et al., 2001; Troyer et al., 1998; Eguchi et al., 2018). See Appendix II for detailed Python code.

2.2 Object-like Visual Stimuli

I designed a set of intrinsically structured, object-like visual stimuli to investigate feature binding in the in silico cortex. Each stimulus consisted of a central body and several bilaterally symmetrical limbs (Fig 2.3a), with its shape defined by three orthogonal feature dimensions (Fig 2.3b):

- *Thorax length*, the distance between the branching points of front and hind limbs
- *Arm length*, the extension distances of middle limbs
- *Leg angle*, the branching angle of the front and hind limbs

This unique shape configuration encouraged the cortex to learn the abstract geometric relationships among different parts of an object without localised spatial cues, a core challenge of the binding problem. I generated each stimulus as a 32×32-pixel binary image with a 2-pixel stroke width. To avoid undesired pixel overlaps, three representative values along each feature dimension were selected to produce a set of 27 object-like visual stimuli for subsequent training and testing (Fig 2.4). All stimuli were generated using NumPy v2.2.0 in Python 3.12.

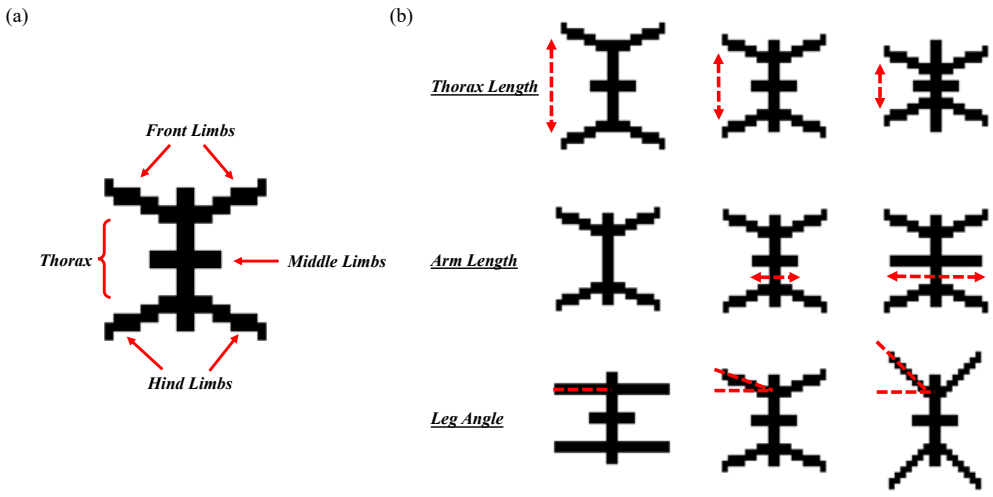

Fig 2.3: (a) Overall shape of the designed object-like visual stimulus and (b) its three underlying feature dimensions.

Fig 2.4: The resulting 27-object stimulus set. The selected values from each dimension: [*16, 11, 5 pixels*] for thorax length, [*0, 8, 16 pixels*] for arm length, and [*0°, 22.5°, 45°*] for leg angle.

2.3 Simulation Execution

I built the in silico visual cortex using the Numba CUDA v0.61.0 (Lam et al., 2015) in Python 3.12 and executed it on GPU-accelerated (NVIDIA A100-SXM) clusters provided by the Cambridge Center for Data Driven Discovery. Before training, excitatory-excitatory synaptic efficacies were randomised by sampling uniformly from [0, 1], and all inhibitory synaptic efficacies were fixed at 1.

I trained the cortex on a total of 5,400 stimuli, each randomly selected, with replacement, from the set of 27 object-like visual stimuli. Each stimulus was displayed at a randomised retinal location for 2,000 ms, during which the cortex was updated according to the aforementioned equations, with differential equations solved discretely using the forward Euler method with a timestep of $\delta t = 0.1$ ms. After each image presentation, all neuronal and synaptic activities were reset to their resting states, except the updated synaptic efficacies.

I evaluated neurons in the cortex over eleven testing sessions: one before training, one after training, and nine intermediate sessions during training (once per 540 stimuli trained). Each visual stimulus in the stimulus set was displayed at the center of retina for 1,000 ms with

synaptic plasticity disabled, during which the firing rates of layer 4 excitatory neurons, the cortex's output units, were recorded for subsequent analyses.

2.4 Neural Tuning Analysis

I applied information theory to characterise neurons' tuning properties and assess the formation of 2-gram neurons. All tuning analyses were conducted using NumPy v2.2.0 and SciPy v1.15 in Python 3.12.

Following Luczak (2024), a neuron's selectivity for a given feature dimension k (e.g., *leg angle*) was quantified via the information entropy, H_k, carried by its firing responses:

$$H_k = -\sum_{n \in k} P(r|n) \cdot \log_2 P(r|n) \tag{11}$$

where n denotes a specific feature value along k (e.g., *45°*), and $P(r|n)$ is the conditional probability of firing (i.e., the neuron's normalised average firing rate) given that feature n is displayed. A neuron more selective for k should exhibit a more unevenly distributed firing rate along k, thus yielding a lower H_k. The neuron's normalised feature selectivity, λ_k, was then computed via:

$$\lambda_k = 1 - \frac{\lambda_k}{\log_2 \|k\|} \tag{12}$$

where $\|k\| = 3$ is the number of distinct feature values along k. Since $\log_2 \|k\|$ is the theoretical upper bound of H_k, λ_k ranges from 0 (no selectivity) to 1 (maximum selectivity).

To quantify the number of feature dimensions to which a neuron is selective for, I defined a neuron's *effective dimensionality*, D_{eff}, as the scale-independent participation ratio (Recanatesi et al., 2022) of its normalised feature selectivity across the three orthogonal feature dimensions ($k \in \{1, 2, 3\}$):

$$D_{\text{eff}} = \frac{(\sum_{k=1}^{3} \lambda_k)^2}{\sum_{k=1}^{3} (\lambda_k)^2} \tag{13}$$

To measure a neuron's absolute sensitivity to its preferred feature values, I defined a neuron's *tuning strength*, ε, as its largest normalised feature selectivity across the three feature dimensions:

$$\varepsilon = \max_k(\lambda_k) \tag{14}$$

By definition, D_{eff} ranges from 1 (selective for a single dimension) to 3 (selective for all three dimensions) and ε ranges from 0 (zero sensitivity) to 1 (maximum sensitivity). Thus, the expected tuning profiles of ideal neurons were:

- 2-gram neurons: $\mathrm{D_{eff}} = 2$ and $\varepsilon = 1$
- Neurons tuned to a single feature: $\mathrm{D_{eff}} = 1$ and $\varepsilon = 1$
- Neurons tuned to an entire stimulus: $\mathrm{D_{eff}} = 3$ and $\varepsilon = 1$
- Untuned neurons: $\varepsilon = 0$

To categorise neurons in the cortex, I adopted the following classification boundaries:

- *2-gram neurons*: $1.5 < D_{\text{eff}} \leq 2.5$ and $\varepsilon > 0.2$
- *Non-2-gram neurons*: ($D_{\text{eff}} \leq 1.5$ or $D_{\text{eff}} > 2.5$) and $\varepsilon > 0.2$

- *Sharply tuned neurons*: $\varepsilon > 0.2$
- *Untuned neurons*: $\varepsilon \leq 0.2$

2.5 Positional Invariance Analysis

I analysed a neuron's positional invariance by displaying each visual stimulus in the stimulus set for 1,000 ms at eight selected symmetrically offset locations, with synaptic plasticity disabled (Fig 2.5). The neuron's firing responses for different stimuli across locations were compared to evaluate its positional invariance.

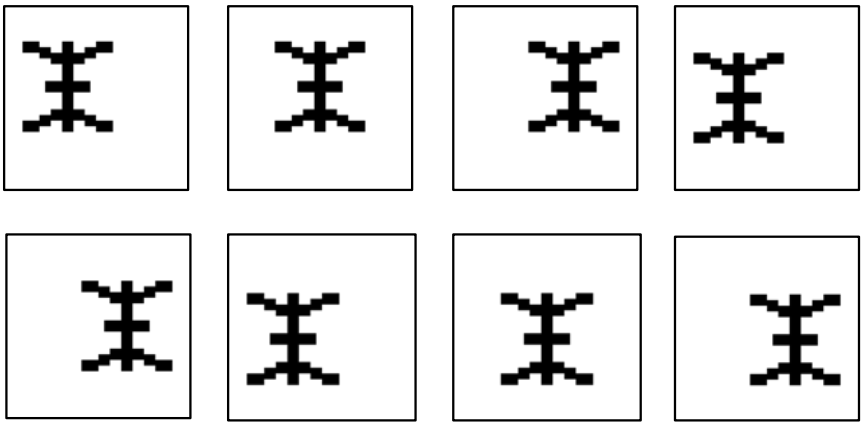

Fig 2.5: Eight retinal offset locations selected for analysing positional invariance. Vertical offsets: *±2 pixels*; horizontal offsets: *±5 pixels*.

2.6 Representational Similarity Analysis

I applied representational similarity analysis (Kriegeskorte et al., 2008) to evaluate whether a neural population explicitly encodes different feature combinations in object-like visual stimuli. I first computed the pairwise representational dissimilarities among visual stimuli and among their elicited neural responses. For a pair of stimuli $\boldsymbol{f_i}$ and $\boldsymbol{f_j}$, its stimulus dissimilarity, S_{ij}, was defined as:

$$S_{ij} = \frac{\mathrm{H}_d(\boldsymbol{f_i}, \boldsymbol{f_j})}{3} \tag{15}$$

where H_d, the Hamming distance, counts the number of unshared feature values, which ranges from 1 (one feature unshared) to 3 (all features unshared). Similarly, for a pair of neural responses $\boldsymbol{r_i}$ and $\boldsymbol{r_j}$, its neural dissimilarity, R_{ij}, was defined as:

$$R_{ij} = \frac{1 - \mathrm{Corr}(\boldsymbol{r_i}, \boldsymbol{r_j})}{2} \tag{16}$$

where $\mathrm{Corr}(\boldsymbol{r_i}, \boldsymbol{r_j})$ indicates Pearson correlation coefficient between $\boldsymbol{r_i}$ and $\boldsymbol{r_j}$. All S_{ij} and R_{ij} ranged from 0 to 1.

I then assembled all S_{ij} and R_{ij} values into representational dissimilarity matrices (RDMs), which were 27×27 symmetrical matrices with each row and column indexing a specific visual stimulus or its elicited neural response. The embedding spaces of the visual stimuli or the neural responses were visualised by projecting their RDMs onto two-dimensional maps using the SMACOF algorithm (Kruskal, 1964) with 10 different initialisations and 300 maximum iterations. The similarity score between the stimulus set and the targeted neural population were quantified by Spearman's rank-order correlation across the off-diagonal entries of their RDMs.

All representational similarity analyses were conducted using NumPy v2.2.0 in Python 3.12, with the SMACOF algorithm applied using scikit-learn v1.16 in Python 3.12.

2.7 Statistical Analysis

I tested hypotheses regarding pairwise group-level differences in mean using one-way repeated measures ANOVA with Bonferroni's *post hoc* comparisons. Assumptions for ANOVA were first verified[1], and if violated, nonparametric Friedman's test with Conover's *post hoc* comparisons was conducted instead. Hypotheses regarding correlations were tested using Spearman rank-based correlation analysis. A significant level of $\alpha = 0.05$ was applied to all statistical tests. All statistical analyses were conducted in JASP v0.18.3 (JASP Team, 2019).

[1] Assumption of normality tested by Shapiro-Wilk test, homogeneity of variance tested by Levene's test, and sphericity tested by Mauchly's test.

3. Results

3.1 The Emergence of 2-gram Neurons

Do 2-gram neurons emerge spontaneously in the visual cortex? I trained the in silico visual cortex by exposing it to 5,400 object-like visual stimuli. Over the course of training, synaptic efficacies of excitatory-excitatory connections rapidly evolved from a uniform distribution ($M = 0.501$, $SD = 0.289$) to a symmetrical peak ($M = 0.503$, $SD = 0.210$), with both its mean and standard deviation roughly converged to stable values, albeit with fluctuations, indicating a reasonably well-behaved learning process (Fig 3.1).

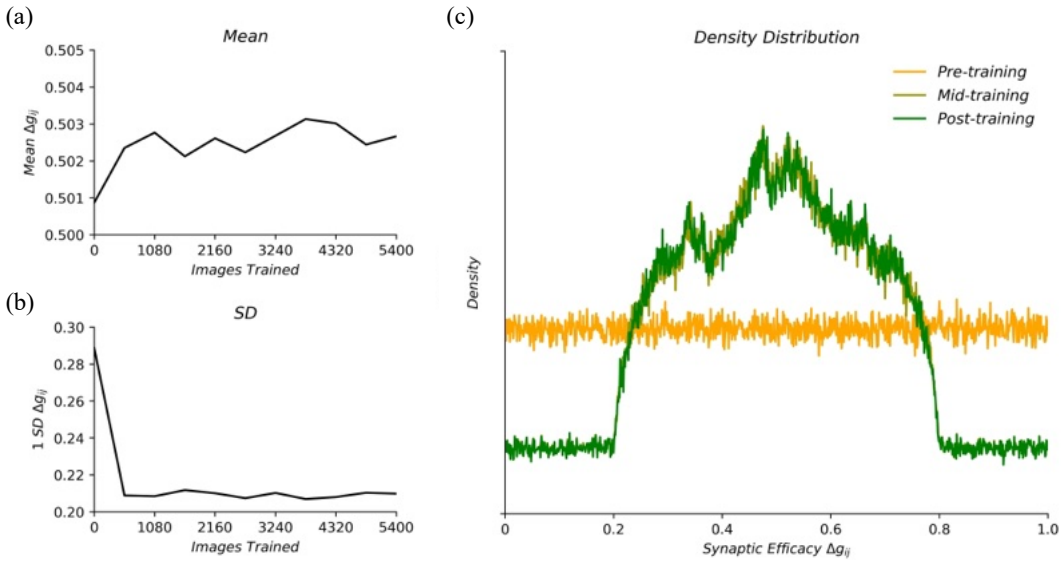

Fig 3.1: Variations in excitatory-excitatory synaptic efficacies Δg_{ij} across training, including their (a) mean, (b) standard deviation, and (c) distribution. *Pre-training* = 0 images trained. *Mid-training* = 2700 images trained. *Post-training* = 5,400 images trained.

To examine whether 2-gram-like tuning properties emerged through the cortex's intrinsic plasticity, I analysed neurons' effective dimensionality D_{eff} and tuning strength ε before, midway, and after training. As training progressed, the distribution of D_{eff} shifted from a right-skewed curve to a concentrated peak around $D_{\text{eff}} \approx 2.3$, approaching the expected $D_{\text{eff}} = 2.0$

for ideal 2-gram neurons (Fig 3.2). Consistent with this observation, the population mean of D_{eff} varied significantly across training, despite with a small effect size (Friedman test: $\chi^2(2) = 60.8$, $p < .001$, Kendall's $W = 0.03$), which decreased non-significantly from pre- to mid-training and then further decreased significantly by post-training (Table 1a). The distribution of ε, in contrast, remained heavily skew towards $\varepsilon = 0$ throughout training, far from the expected $\varepsilon = 1$ for sharply tuned neurons (Fig 3.2b). Although significant variations in the population mean of ε were detected (Friedman test: $\chi^2(2) = 228.8$, $p < .001$, Kendall's $W = 0.11$), it first decreased from pre- to mid-training and then increased by post-training, resulting in a nonsignificant net change across training (Table 1b). This consistently low ε was likely due to the dense lateral inhibitory connections within each cortical layer, which prevents the formation of sharply tuned neurons. Together, these findings reveal the emergence of population-wide, low-strength, 2-gram-like tuning in the cortex.

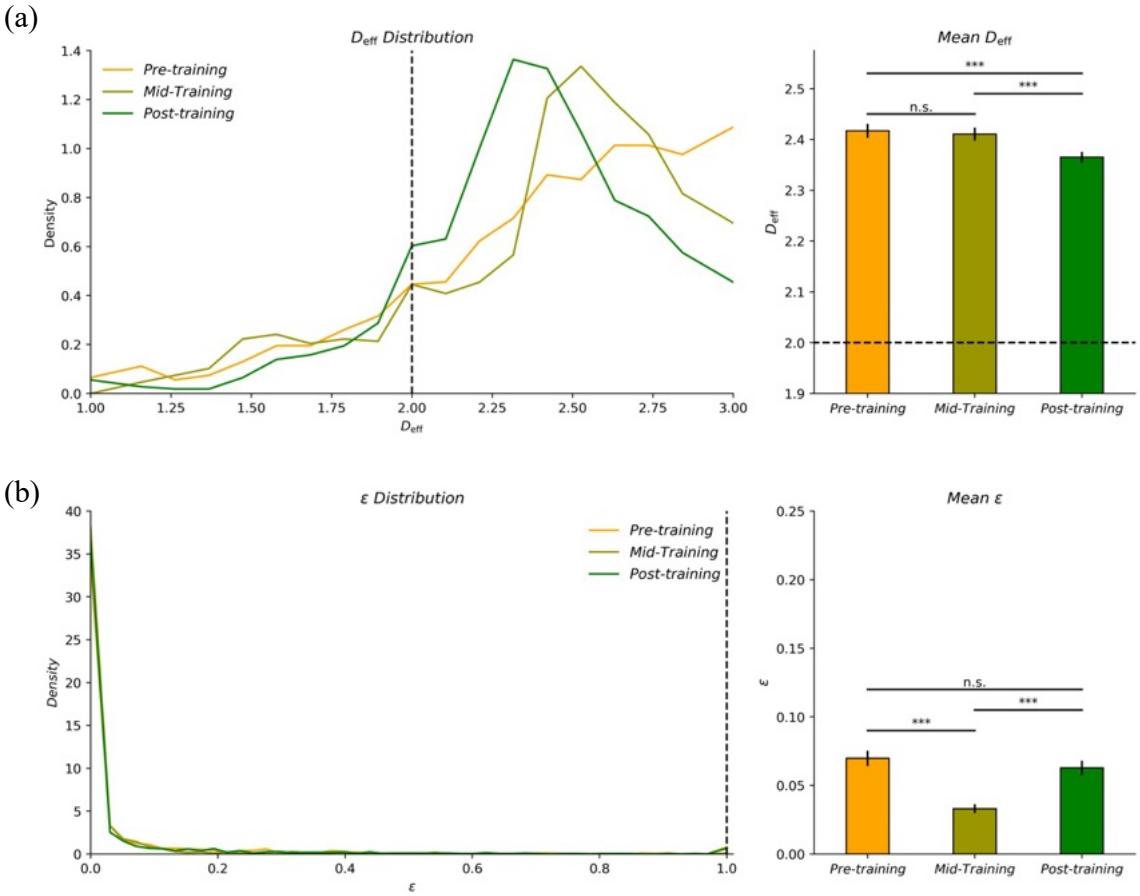

Fig 3.2: Distributions and population means of (a) effective dimensionality D_{eff} and (b) tuning strength ε for neurons before, midway, and after training. Distributions estimated using 20 evenly spaced histogram-bins. *Pre-training*: 0 images trained. *Mid-training*: 2700 images trained. *Post-training*: 5,400 images trained. Dotted lines indicate the expected D_{eff} and ε for an ideal 2-gram neural population. Error bars indicate $\pm$ 1 SEM. *** represents $p < .001$.

Table 1a: Pairwise comparisons of D_{eff} before, midway, and after the training.

Condition	*Summary*		*Pairwise vs. Pre-training*		*Pairwise vs. Mid-training*	
	M	***SD***	***t***	p_{bonf}	***t***	p_{bonf}
Pre-training	2.43	0.44	/	/	/	/
Mid-training	2.41	0.41	1.91	0.17	/	/
Post-training	2.36	0.35	7.50	*<.001*	5.59	*<.001*

Pairwise results from Conover's *post hoc* comparisons. p_{bonf}: p-value after Bonferroni correction.

Table 1b: Pairwise comparisons of ε before, midway, and after the training.

Condition	*Summary*		*Pairwise vs. Pre-training*		*Pairwise vs. Mid-training*	
	M	***SD***	***t***	p_{bonf}	***t***	p_{bonf}
Pre-training	0.07	0.18	/	/	/	/
Mid-training	0.03	0.10	13.1	*<.001*	/	/
Post-training	0.06	0.17	0.01	1.00	13.1	*<.001*

Pairwise results from Conover's *post hoc* comparisons. p_{bonf}: p-value after Bonferroni correction.

To further investigate the emergence of sharply tuned 2-gram neurons in the cortex, I assessed each neuron's tuning profile during training across eleven evenly spaced testing sessions. As shown in Fig 3.3, most neurons remained untuned throughout training, but a subset of neurons acquired high ε values and gradually clustered between $D_{\text{eff}} \approx 2.0$ and 3.0. I then classified these sharply tuned neurons as 2-gram or non-2-gram based on their D_{eff} values and tracked their population sizes as training progressed (Table 2). Both populations initially declined from $n \approx 50$ to $n \approx 20$, likely due to an early synaptic pruning. Subsequently, the 2-gram population rebounded and peaked at $n = 71$, while the non-2-gram population oscillated between $n = 16$ and $n = 61$ (Fig 3.4a). As a result, while the total number of sharply tuned neurons in the cortex

remained roughly unchanged at $n \approx 100$ throughout training, the proportion of 2-gram neurons among them increased from 51% to 58% across training, despite with large fluctuations (Fig 3.4b). Together, these findings demonstrate the gradual, yet unstable, emergence of sharply tuned 2-gram neurons in the cortex.

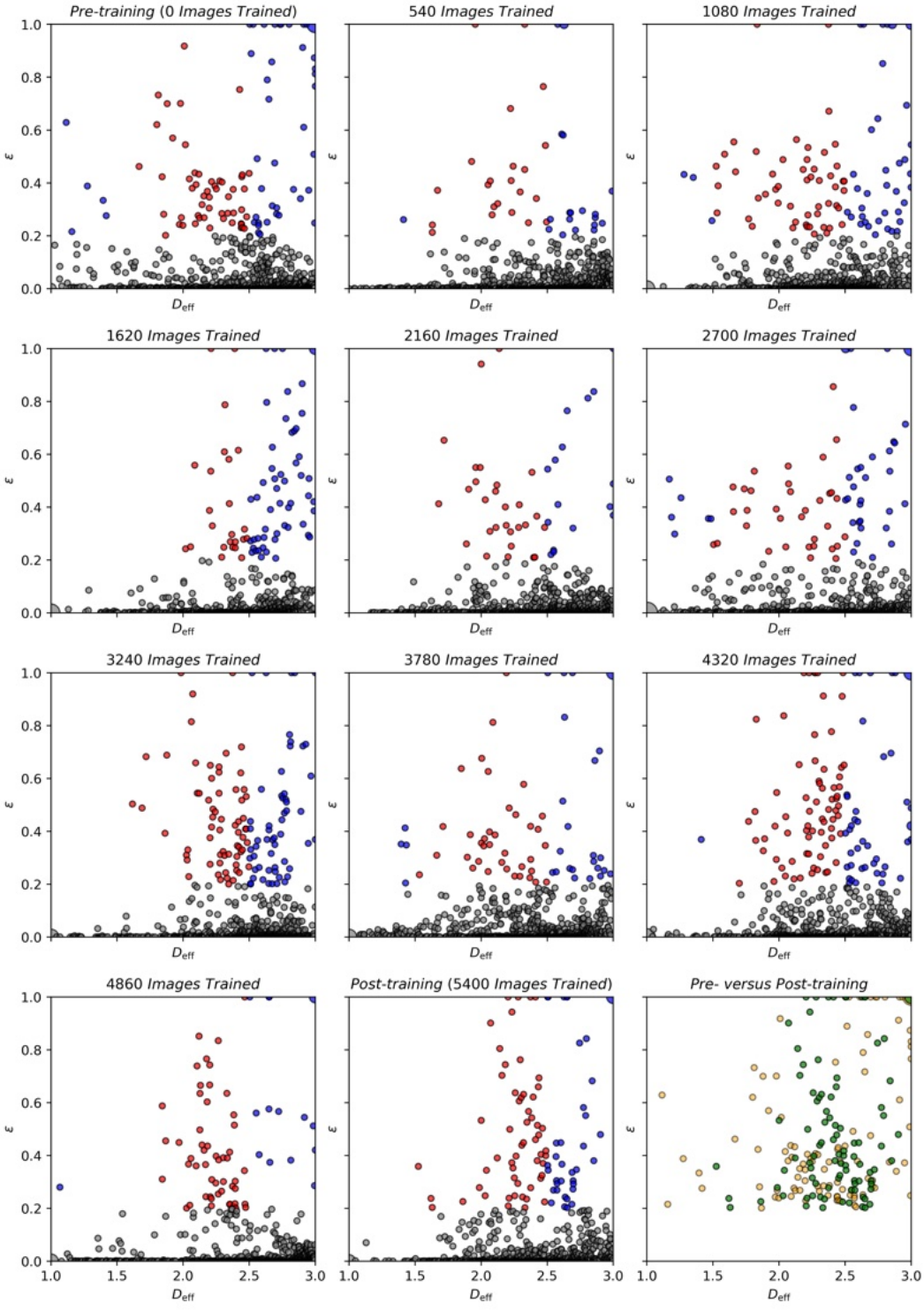

Fig 3.3: Neurons' tuning profiles during training measured across eleven testing sessions. Red, blue, and gray datapoints respectively indicate neurons classified as sharply tuned 2-gram, sharply tuned non-2-gram, and untuned neurons. In the bottom right subplot, orange and green datapoints respectively indicate all sharply tuned neurons pre- and post-training. Marker sizes reflect the number of overlapping neurons that shared the same tuning profile.

Table 2: Number and relative proportion of sharply tuned 2-gram, sharply tuned non-2-gram, and untuned neurons during training.

Images Trained	*2-gram Neurons*		*Non-2-gram Neurons*		*Untuned Neurons*	
	n	***%***	***n***	***%***	***n***	***%***
0	53	5.12	51	4.98	920	89.8
540	22	2.15	20	1.95	982	95.9
1080	52	5.78	43	4.20	929	90.7
1620	24	2.34	55	5.37	945	92.3
2160	28	2.73	16	1.56	980	95.7
2700	34	3.32	43	4.20	947	92.5
3240	65	6.35	56	5.47	903	88.2
3780	37	3.61	31	3.03	956	93.4
4320	71	6.93	48	4.69	905	88.4
4860	52	5.01	18	1.76	954	93.2
5400	56	5.47	40	3.91	928	90.3

n: Raw counts. ***%***: Proportion among all neurons ($n = 1024$) across the cortical layer.

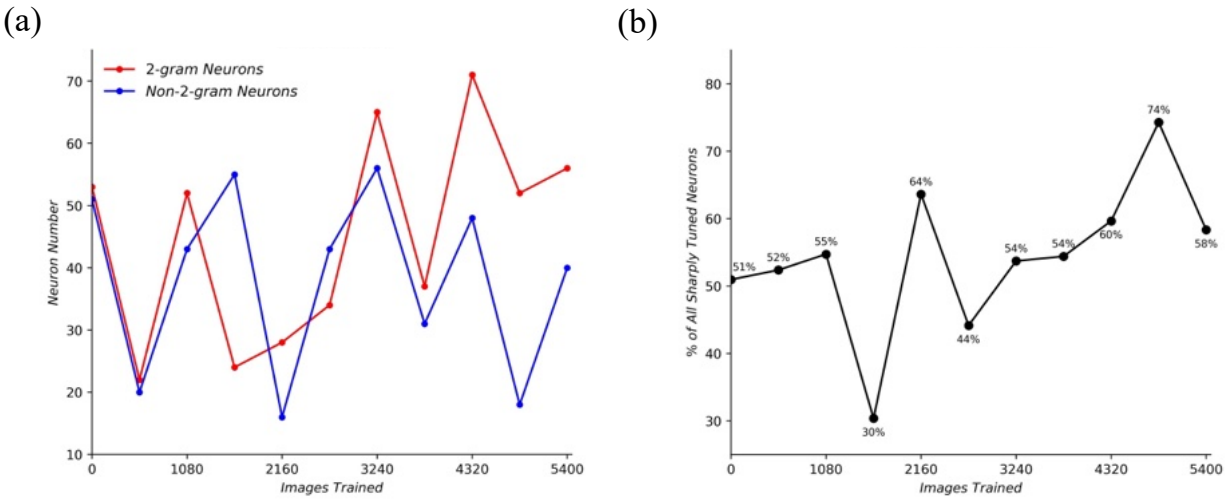

Fig 3.4: (a) Absolute size and (b) relative proportion of the 2-gram neural population among all sharply tuned neurons as training progressed.

3.2 Positional Invariance of 2-gram Neurons

Do the emerged 2-gram neurons respond invariantly to visual stimuli across retinal locations? After training the cortex, I presented each object-like visual stimulus at eight retinal offsets and tracked how the tuning profiles of previously identified 2-gram neurons varied. All targeted 2-gram neurons exhibited great fluctuations in their tuning properties (Fig 3.5 and Table 3), as their population mean D_{eff} and ε differed significantly across locations (Friedman test: $\chi^2(32) > 60.3$, $p < .001$). Further pairwise comparisons revealed that D_{eff} was significantly lower than the central value at four offsets (Conover's *post hoc* comparisons: $t > 3.168$, $p_{bonf} < .05$) and ε was significantly lower at all eight offsets ($t > 3.38$, $p_{bonf} < .03$; Table 4). Out of the $n = 56$ sharply tuned 2-gram neurons identified at the central location, $n = 10$ retained their 2-gram properties at one offset, $n = 1$ at two offsets, and none beyond two offsets. Overall, these results indicate that the emerged 2-gram neurons are not positionally invariant.

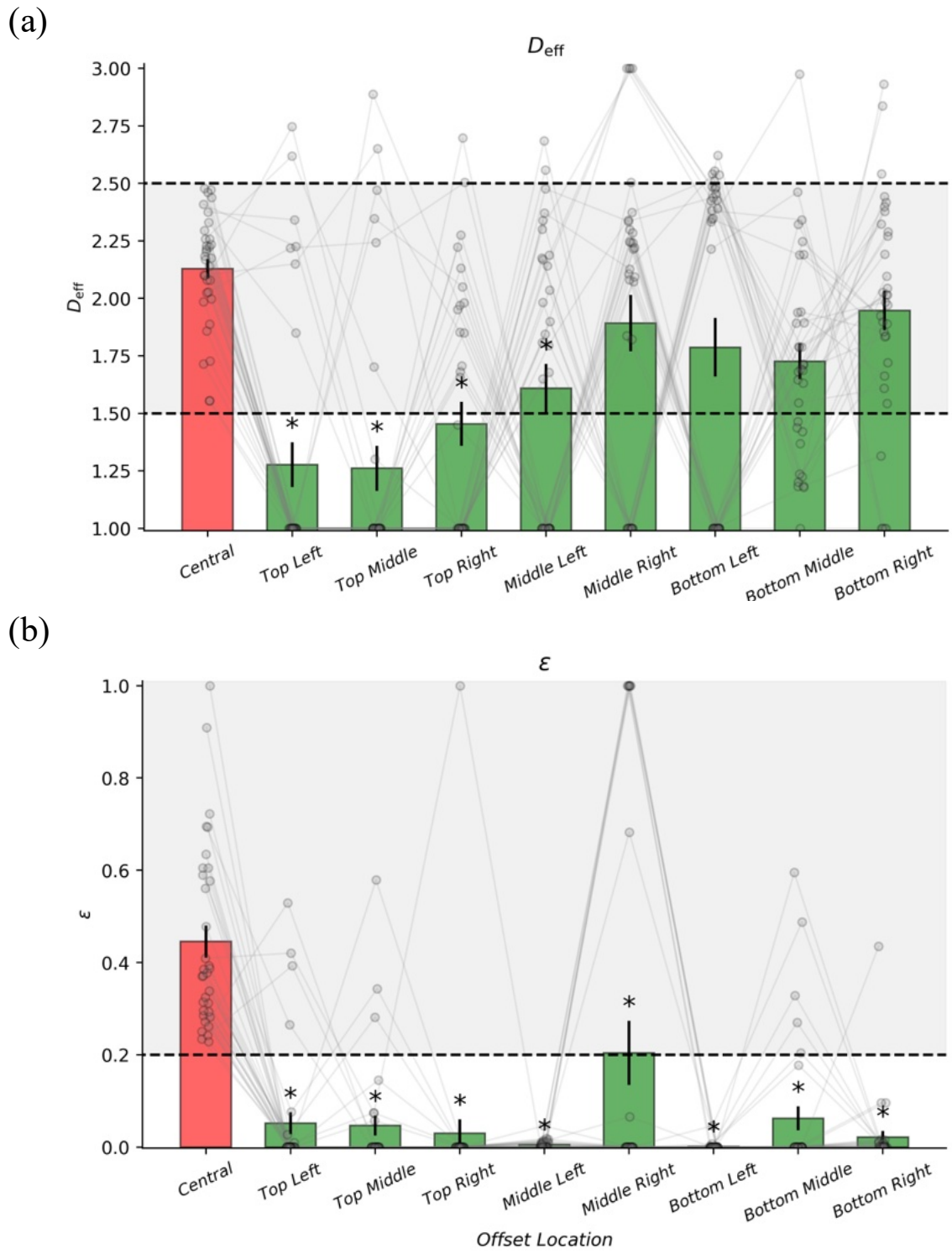

Fig 3.5: Variations in (a) $D_{\mathbf{eff}}$ and (b) $\boldsymbol{\varepsilon}$ of centrally classified 2-gram neurons across eight offset retinal locations. Datapoints represent individual neurons, with lines connecting the same neurons across conditions. Gray areas between dotted lines indicate the classification boundaries for sharply tuned 2-gram neurons. Error bars indicate $\pm$ 1 SEM. * represents significant ($p < .05$) pairwise difference versus the central value.

Table 3: Tuning properties of centrally identified 2-gram neurons across retinal offsets.

Offset Location	D_{eff}		ε		n
	M	***SD***	***M***	***SD***	
Central	*2.13*	*0.25*	*0.45*	*0.20*	*56*
Top Left	1.28	0.56	0.05	0.14	4
Top Middle	1.26	0.56	0.05	0.13	1
Top Right	1.45	0.55	0.03	0.17	0
Middle Left	1.61	0.61	0.01	0.01	0
Middle Right	1.89	0.70	0.21	0.40	2
Bottom Left	1.79	0.73	0.00	0.01	0
Bottom Middle	1.73	0.44	0.06	0.15	4
Bottom Right	1.95	0.50	0.02	0.02	1

n: Counts of neurons that retained 2-gram tuning properties at the specified offset.

Table 4: Pairwise comparisons of D_{eff} and ε at different retinal offsets versus the central value.

Offset Location	D_{eff}		ε	
	t	p_{bonf}	***t***	p_{bonf}
Top Left	5.32	*<.001*	8.04	*<.001*
Top Middle	5.05	*<.001*	7.99	*<.001*
Top Right	3.83	*0.01*	6.57	*<.001*
Middle Left	3.17	*0.05*	6.15	*<.001*
Middle Right	0.90	1.00	4.78	*<.001*
Bottom Left	1.79	1.00	7.53	*<.001*
Bottom Middle	1.95	1.00	3.38	*0.03*
Bottom Right	0.71	1.00	4.73	*<.001*

Pairwise results from Conover's *post hoc* comparisons. p_{bonf}: *p*-value after Bonferroni correction.

3.3 Feature Representations among 2-gram Neurons

Can the emerged 2-gram neurons explicitly encode feature combinations in the object-like visual stimuli? To address this question, I applied representational similarity analysis to assess whether 2-gram neural responses directly reflect the stimulus set's underlying feature space.

To qualitatively compare the representational similarities between visual stimuli and 2-gram neural responses, I computed the pairwise stimulus and 2-gram neural dissimilarities, constructed their respective RDMs, and projected them onto two-dimensional embedding spaces via multidimensional scaling. While the stimulus representations formed evenly distributed subgroups, reflecting the continuous variation in their underlying feature combinations (Fig 3.6a), the corresponding 2-gram representations formed tightly spaced clusters with no clear grouping pattern (Fig 3.6b). This discrepancy indicates a substantial mismatch between the stimulus set and 2-gram neural responses in representational geometries.

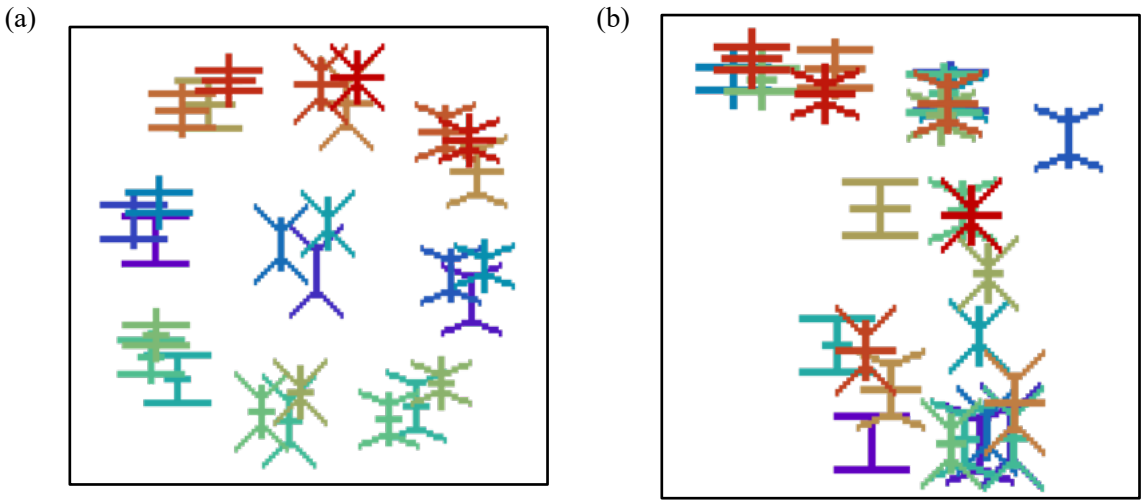

Fig 3.6: Two-dimensional embedding spaces of (a) visual stimuli and (b) their elicited 2-gram neural responses. Falsely coloured items indicate stimulus representations in (a) and neural representations in (b), with distances between items reflect pairwise dissimilarities.

To further assess the representational similarity between the stimulus set and 2-gram neural responses, I computed the rank-based similarity score between the stimulus RDM (Fig 3.7a) and the 2-gram neural RDM (Fig 3.7b). For comparison, I also included RDMs for non-2-gram neurons, untuned neurons, and the combined whole cortex in the analysis. All neural RDMs exhibited statistically significant yet modest correlations with the stimulus RDM (Spearman's

rank-order correlation: $\rho = 0.16$ to 0.29, $p < .003$; Fig 3.7c and Table 5), indicating that each neural population in the cortex shares, to some extent, representational similarity with the stimulus set. Notably, the 2-gram neural RDM exhibited the highest similarity score with the stimulus RDM, while its counterpart, the non-2-gram neural RDM, exhibited the lowest similarity score, suggesting that 2-gram neural responses align most closely with the stimulus set's underlying feature space. However, even this highest score is still low in absolute magnitude and only marginally exceeds that of the other neural populations. Combined with the qualitative mismatch in the embedding space, all results indicate that the emerged 2-gram neurons do not explicitly encode feature combinations in visual stimuli.

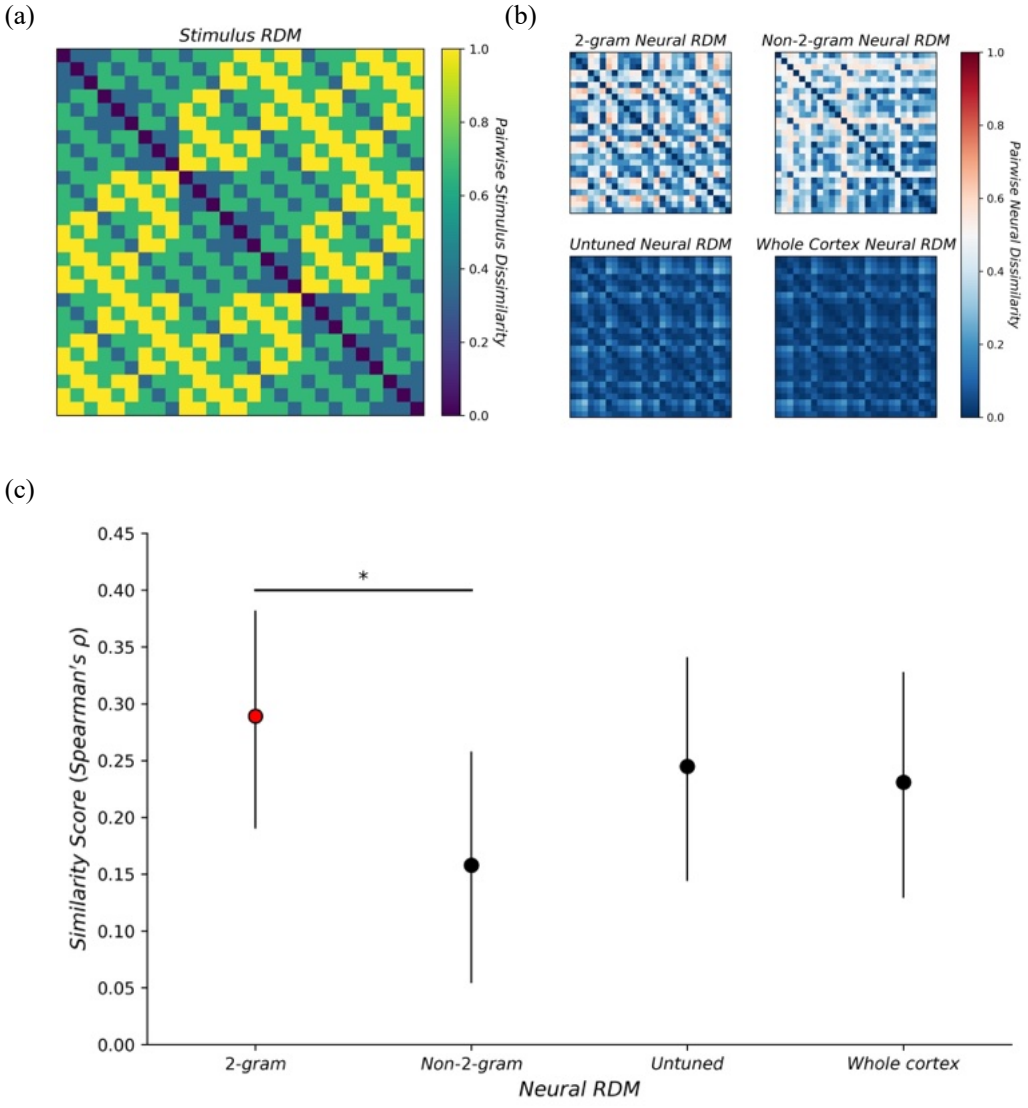

Fig 3.7: (a) Stimulus and (b) neural RDMs constructed, with (c) the similarity score (Spearman correlation coefficient) between the stimulus RDM and each neural RDM. Error bars indicate 95% confidence intervals. * represents pairs with non-overlapping confidence intervals.

Table 5: Representational similarity score between the stimulus RDM and each neural RDM.

Neural RDM	*Spearman's ρ*	*95% Confidence Interval*	*p*
2-gram	0.29	[0.19, 0.38]	*<.001*
Non-2-gram	0.16	[0.05, 0.26]	*<.003*
Untuned	0.25	[0.14, 0.34]	*<.001*
Whole Cortex	0.23	[0.13, 0.33]	*<.001*

4. Discussion

4.1 Conclusions

In this study, I investigated the emergence of 2-gram neurons in a spiking neural network model of the human visual cortex and assessed their potential as a novel mechanism for addressing the FBP. By presenting the cortex with object-like visual stimuli and updating it via unsupervised learning, I showed the gradual emergence of 2-gram neurons in the cortex, specifically, both low-amplitude, 2-gram-like tuning at population level and unstable, sharply tuned 2-gram neurons at individual levels. By probing the cortex at various retinal locations, I revealed that these neurons are not positionally invariant. By decoding their firing patterns, I further demonstrated that the emerged 2-gram neurons do not explicitly encode feature combinations in visual stimuli. Together, this study demonstrates the spontaneous emergence of 2-gram neurons in a neurobiologically realistic model of the human visual system and elucidates their potential and limitation as a novel solution to the FBP.

A surprising property of the emerged 2-gram neurons in the in silico cortex is their instability. Even after training on 5,400 images, the tuning properties of 2-gram neurons still had not stabilised. While this instability most likely reflects the incomplete convergence of the model, an alternative interpretation is that the emerged 2-gram neurons functionally act as subplate neurons in the developing visual cortex (Kanold, 2004), which play transient but critical roles in modulating synaptic weights. Under this idea, the emerged 2-gram neurons are not binding units, but rather transient byproducts of spike-timing dependent plasticity in the cortex. This alternative interpretation explains why all sharply tuned 2-gram neurons essentially lack key properties for feature binding. However, this alternative explanation should be taken with caution, as it might be an overinterpretation.

The findings of the present study, to some extent, confirm previous works that support the involvement of 2-gram-like units in feature binding (Morita et al., 2010; Furutate et al., 2019; Schneegans & Bays, 2017). Specifically, I demonstrated that neurons tuned to pairwise feature conjunctions can emerge spontaneously through repetitive, unsupervised sensory exposures in a hierarchical neural network. This result provides a neurobiological foundation for models that use arbitrarily constructed 2-gram units to bind features (Morita et al., 2010; Schneegans & Bays, 2017), as such units can naturally form without the need of innate hardwiring mechanisms or supervised learning rules. However, unlike idealised 2-gram units proposed in these abstract models, the emerged 2-gram neurons only exhibited weak if any, ability to bind features. This lack of binding capability may reflect methodological constraints, as discussed below, or may indicate that 2-gram neurons need complementary active mechanisms to bind features reliably. Indeed, a complete solution to the FBP likely involves both a passive neural architecture, which defines what to bind, and an active neural mechanism, which specifies how to bind. Future work should therefore explore how the emerged 2-gram neurons interact with popular active binding mechanisms, such as binding by synchrony (von der Malsburg, 1981) and binding by rate enhancement (Treisman & Gelade, 1980; Reynolds & Desimone, 1999) to truly assess their capacity for feature binding.

4.2 Limitations and future directions

This study has several limitations that future works should address. First, the in silico cortex that I constructed relies exclusively on bottom-up visual inputs and omits top-down neural influences (Gilbert & Li, 2013), notably top-down attentional modulation (Noudoost et al., 2010), which is important for feature binding in a biological visual cortex (Roelfsema, 2023). Similarly, for runtime considerations, I ignored synaptic delays and neural noise in the cortex, which are known to influence the stability and performance of spiking neural networks (Ma et

al., 2023). Future works should incorporate these missing neural processes, namely, top-down attentional modulation, synaptic delays, and neural noise, to construct a more biologically realistic in silico visual cortex for investigating feature binding.

Second, I updated all synaptic weights in the cortex by a local unsupervised learning rule. While unsupervised learnings are powerful and reflect key aspects of perceptual learning in human (Sumner, 2020), it can also induce network instability (van Hemmen, 1997). Incorporating supervised or reinforcement learning with explicit objectives, for example, via multi-object classification or discrimination tasks, could in theory stabalise the cortex and encourage the occurrence positionally invariant neurons with strong binding abilities. Future works should compare and contrast the effect of different training protocols the formation of 2-grams or other binding-related neurons in the cortex.

Finally, I applied entropy-based analysis on measuring each neuron's tuning profile, where I effectively projected each neuron's activities from a three-dimensional space onto a single axis for computation. Such approach ignored useful information regarding correlations between multiple feature dimensions, temporal dynamics of neural firing patterns, and covariance structures between multiple neurons. Future works should use more sophisticated analytic tools, such as topological data analysis (Lin & Kriegeskorte, 2024), to examine the tuning properties of neurons in the cortex with less constrains.

Reference

Barlow, H. B. (1972). Single Units and Sensation: A Neuron Doctrine for Perceptual Psychology? *Perception*, *1*(4), 371–394. https://doi.org/10.1068/p010371

Burkitt, A. N. (2006). A Review of the Integrate-and-fire Neuron Model: I. Homogeneous Synaptic Input. *Biological Cybernetics*, *95*(1), 1–19. https://doi.org/10.1007/s00422-006-0068-6

Cavnar, W. B., & Trenkle, J. M. (1994, April). N-gram-based text categorization. In *Proceedings of SDAIR-94, 3rd annual symposium on document analysis and information retrieval* (Vol. 161175, p. 14).

Di Lollo, V. (2012). The feature-binding problem is an ill-posed problem. *Trends in Cognitive Sciences*, *16*(6), 317–321. https://doi.org/10.1016/j.tics.2012.04.007

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences*, *11*(8), 333–341. https://doi.org/10.1016/j.tics.2007.06.010

Eguchi, A., Isbister, J. B., Ahmad, N., & Stringer, S. (2018). The emergence of polychronization and feature binding in a spiking neural network model of the primate ventral visual system. *Psychological Review*, *125*(4), 545–571. https://doi.org/10.1037/rev0000103

Fan, X., & Markram, H. (2019). A Brief History of Simulation Neuroscience. *Frontiers in Neuroinformatics*, *13*. https://doi.org/10.3389/fninf.2019.00032

Feldman, J. (2013). The neural binding problem(s). *Cognitive Neurodynamics*, *7*(1), 1–11. https://doi.org/10.1007/s11571-012-9219-8

Fiser, J., & Lengyel, G. (2022). Statistical Learning in Vision. *Annual Review of Vision Science*, *8*(Volume 8, 2022), 265–290. https://doi.org/10.1146/annurev-vision-100720-103343

Fries, P., Nikolić, D., & Singer, W. (2007). The gamma cycle. *Trends in Neurosciences*, *30*(7), 309–316. https://doi.org/10.1016/j.tins.2007.05.005

Furutate, M., Fujii, Y., Morita, H., & Morita, M. (2019). Visual Feature Integration of Three Attributes in Stimulus-Response Mapping Is Distinct From That of Two. *Frontiers in Neuroscience*, *13*. https://doi.org/10.3389/fnins.2019.00035

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews. Neuroscience*, *14*(5), 10.1038/nrn3476. https://doi.org/10.1038/nrn3476

Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, *117*(4), 500–544. https://doi.org/10.1113/jphysiol.1952.sp004764

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*(1), 106–154. https://doi.org/10.1113/jphysiol.1962.sp006837

Isbister, J. B., Eguchi, A., Ahmad, N., Galeazzi, J. M., Buckley, M. J., & Stringer, S. (2018). A new approach to solving the feature-binding problem in primate vision. *Interface Focus*, *8*(4), 20180021. https://doi.org/10.1098/rsfs.2018.0021

Izhikevich, E. M. (2003). Simple model of spiking neurons. *IEEE Transactions on Neural Networks*, *14*(6), 1569–1572. https://doi.org/10.1109/TNN.2003.820440

J L van Hemmen. (1997). Hebbian learning, its correlation catastrophe, and unlearning. *Network: Computation in Neural Systems*, *8*(3), V1. https://doi.org/10.1088/0954-898X/8/3/001

Jeurissen, D., Self, M. W., & Roelfsema, P. R. (2016). Serial grouping of 2D-image regions with object-based attention in humans. *eLife*, *5*, e14320. https://doi.org/10.7554/eLife.14320

Jolivet, R., Rauch, A., Lüscher, H.-R., & Gerstner, W. (2006). Predicting spike timing of neocortical pyramidal neurons by simple threshold models. *Journal of Computational Neuroscience*, *21*(1), 35–49. https://doi.org/10.1007/s10827-006-7074-5

Kanold, P. O. (2004). Transient microcircuits formed by subplate neurons and their role in functional development of thalamocortical connections. *Neuroreport*, *15*(14), 2149–2153. https://doi.org/10.1097/00001756-200410050-00001

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: An expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, *17*(1), 26–49. https://doi.org/10.1016/j.tics.2012.10.011

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2,* 249. https://doi.org/10.3389/neuro.06.004.2008

Kruizinga, P., & Petkov, N. (1999). Nonlinear operator for oriented texture. *IEEE Transactions on Image Processing*, *8*(10), 1395–1407. https://doi.org/10.1109/83.791965

Kruskal, J. B. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, *29*(2), 115–129. https://doi.org/10.1007/BF02289694

Lam, S. K., Pitrou, A., & Seibert, S. (2015). Numba: A LLVM-based Python JIT compiler. *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*, 1–6. https://doi.org/10.1145/2833157.2833162

Lin, B., & Kriegeskorte, N. (2024). The topology and geometry of neural representations. *Proceedings of the National Academy of Sciences*, *121*(42), e2317881121. https://doi.org/10.1073/pnas.2317881121

Livingstone, M. S., & Hubel, D. H. (1987). Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *Journal of Neuroscience*, *7*(11), 3416–3468. https://doi.org/10.1523/JNEUROSCI.07-11-03416.1987

Luczak, A. (2024). Entropy of Neuronal Spike Patterns. *Entropy*, *26*(11), Article 11. https://doi.org/10.3390/e26110967

Ma, G., Yan, R., & Tang, H. (2023). Exploiting noise as a resource for computation and learning in spiking neural networks. *Patterns*, *4*(10), 100831. https://doi.org/10.1016/j.patter.2023.100831

Maass, W. (1997). Networks of spiking neurons: The third generation of neural network models. *Neural Networks*, *10*(9), 1659–1671. https://doi.org/10.1016/S0893-6080(97)00011-7

Martin, A. B., & von der Heydt, R. (2015). Spike Synchrony Reveals Emergence of Proto-Objects in Visual Cortex. *Journal of Neuroscience*, *35*(17), 6860–6870. https://doi.org/10.1523/JNEUROSCI.3590-14.2015

Mel, B. W. (1997). SEEMORE: Combining Color, Shape, and Texture Histogramming in a Neurally Inspired Approach to Visual Object Recognition. *Neural Computation*, *9*(4), 777–804. https://doi.org/10.1162/neco.1997.9.4.777

Mel, B. W., & Fiser, J. (2000). Minimizing binding errors using learned conjunctive features. *Neural Computation*, *12*(4), 731–762. https://doi.org/10.1162/089976600300015574

Miconi, T., & VanRullen, R. (2010). The Gamma Slideshow: Object-Based Perceptual Cycles in a Model of the Visual Cortex. *Frontiers in Human Neuroscience*, *4,* 205. https://www.frontiersin.org/articles/10.3389/fnhum.2010.00205

Morita, M., Morokami, S., & Morita, H. (2010). Attribute Pair-Based Visual Recognition and Memory. *PLOS ONE*, *5*(3), e9571. https://doi.org/10.1371/journal.pone.0009571

Nassi, J. J., & Callaway, E. M. (2009). Parallel Processing Strategies of the Primate Visual System. *Nature Reviews. Neuroscience*, *10*(5), 360–372. https://doi.org/10.1038/nrn2619

Noudoost, B., Chang, M. H., Steinmetz, N. A., & Moore, T. (2010). TOP-DOWN CONTROL OF VISUAL ATTENTION. *Current Opinion in Neurobiology*, *20*(2), 183–190. https://doi.org/10.1016/j.conb.2010.02.003

Perrinet, L., Delorme, A., Samuelides, M., & Thorpe, S. J. (2001). Networks of integrate-and-fire neuron using rank order coding A: How to implement spike time dependent Hebbian plasticity. *Neurocomputing*, *38–40*, 817–822. https://doi.org/10.1016/S0925-2312(01)00460-X

Recanatesi, S., Bradde, S., Balasubramanian, V., Steinmetz, N. A., & Shea-Brown, E. (2022). A scale-dependent measure of system dimensionality. *Patterns*, *3*(8), 100555. https://doi.org/10.1016/j.patter.2022.100555

Reynolds, J. H., & Desimone, R. (1999). The Role of Neural Mechanisms of Attention in Solving the Binding Problem. *Neuron*, *24*(1), 19–29. https://doi.org/10.1016/S0896-6273(00)80819-3

Riesenhuber, M., & Poggio, T. (1999). Are Cortical Models Really Bound by the "Binding Problem"? *Neuron*, *24*(1), 87–93. https://doi.org/10.1016/S0896-6273(00)80824-7

Roelfsema, P. R. (2023). Solving the binding problem: Assemblies form when neurons enhance their firing rate—they don't need to oscillate or synchronize. *Neuron*, *111*(7), 1003–1019. https://doi.org/10.1016/j.neuron.2023.03.016

Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature*, *395*(6700), 376–381. https://doi.org/10.1038/26475

Rosenblatt, F. (1962). *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. Spartan Books.

Schneegans, S., & Bays, P. M. (2017). Neural Architecture for Feature Binding in Visual Working Memory. *Journal of Neuroscience*, *37*(14), 3913–3925. https://doi.org/10.1523/JNEUROSCI.3493-16.2017

Serre, T. (2013). Hierarchical Models of the Visual System. In D. Jaeger & R. Jung (Eds.), *Encyclopedia of Computational Neuroscience* (pp. 1–12). Springer. https://doi.org/10.1007/978-1-4614-7320-6_345-1

Singer, W., & Gray, C. M. (1995). Visual Feature Integration and the Temporal Correlation Hypothesis. *Annual Review of Neuroscience*, *18*(1), 555–586. https://doi.org/10.1146/annurev.ne.18.030195.003011

Sumner, R. L., Spriggs, M. J., Muthukumaraswamy, S. D., & Kirk, I. J. (2020). The role of Hebbian learning in human perception: A methodological and theoretical review of the human Visual Long-Term Potentiation paradigm. *Neuroscience & Biobehavioral Reviews*, *115*, 220–237. https://doi.org/10.1016/j.neubiorev.2020.03.013

Taylor, J., & Xu, Y. (2022). Representation of color, form, and their conjunction across the human ventral visual pathway. *NeuroImage*, *251*, 118941. https://doi.org/10.1016/j.neuroimage.2022.118941

Treisman, A. (1988). Features and Objects: The Fourteenth Bartlett Memorial Lecture. *The Quarterly Journal of Experimental Psychology Section A*, *40*(2), 201–237. https://doi.org/10.1080/02724988843000104

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, *6*(2), 171–178. https://doi.org/10.1016/S0959-4388(96)80070-5

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Troyer, T. W., Krukowski, A. E., Priebe, N. J., & Miller, K. D. (1998). Contrast-Invariant Orientation Tuning in Cat Visual Cortex: Thalamocortical Input Tuning and Correlation-Based Intracortical Connectivity. *Journal of Neuroscience*, *18*(15), 5908–5927. https://doi.org/10.1523/JNEUROSCI.18-15-05908.1998

Velik, R. (2012). From simple receptors to complex multimodal percepts: A first global picture on the mechanisms involved in perceptual binding. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00259

von der Malsburg, C. (1994). The Correlation Theory of Brain Function. In E. Domany, J. L. Van Hemmen, & K. Schulten (Eds.), *Models of Neural Networks* (pp. 95–119). Springer New York. https://doi.org/10.1007/978-1-4612-4320-5_2

Wolfe, J. M. (2012). The binding problem lives on: Comment on Di Lollo. *Trends in Cognitive Sciences*, *16*(6), 307–308. https://doi.org/10.1016/j.tics.2012.04.013

Yamazaki, K., Vo-Ho, V.-K., Bulsara, D., & Le, N. (2022). Spiking Neural Networks and Their Applications: A Review. *Brain Sciences*, *12*(7), Article 7. https://doi.org/10.3390/brainsci12070863

Appendix I: Key Parameters in the In Silico Visual Cortex

Model Architectures

Parameter Name	*Layer*			
	1	*2*	*3*	*4*
Number of *ex* neurons within each layer	32×32	32×32	32×32	32×32
Number of *inh* neurons within each layer	16×16	16×16	16×16	16×16
Number of feedforward connections per *ex* neuron	30	100	100	/
Number of feedback connections per *ex* neuron	{0, 10}	{0, 10}	{0, 10}	{0, 10}
Number of lateral *ex* to *ex* connection per *ex* neuron	{0, 10}	{0, 10}	{0, 10}	{0, 10}
Number of lateral *ex* to *inh* connection per *ex* neuron	30	30	30	30
Number of lateral *inh* to *ex* connection per *inh* neuron	30	30	30	30
Projection radius of feedforward connections	0.5	4.0	6.0	8.0
Projection radius of feedback connections	4.0	4.0	4.0	/
Projection radius of lateral *ex* to *ex* connections	2.0	2.0	2.0	2.0
Projection radius of lateral *ex* to *inh* connection	0.5	0.5	0.5	0.5
Projection radius of lateral *inh* to *ex* connections	4.0	4.0	4.0	4.0

Parameters values adopted from Eguchi et al., (2018); *ex*: excitatory neuron; *inh*: inhibitory neuron; {0, 10}: randomly drawn from [0, 10].

Neural and Synaptic Parameters

Parameter Name		*Adopted Value*	*Source*
Simple Cells (Eq 1 - 3)	θ	0, $\frac{\pi}{4}$, $\frac{\pi}{2}$, $\frac{3\pi}{4}$	**c**
	γ	0.5	**c**
	$\bar{\lambda}$	2	**c**
	b	1.5	**c**
	ψ	π	**c**
	ω	100 Hz	**c**
Neurons (Eq 4 - 7)	τ_m^γ	*ex*: 20 ms *inh*:12 ms	**a**
	V_0^γ	*ex*: -74 mV *inh*: -82 mV	**a**
	R^γ	*ex*: 40 MΩ *inh*: 55 MΩ	**a**
	Θ^γ	*ex*: -53 mV *inh*: -53 mV	**a**
	τ_R	20 ms	**a**
	V_H^γ	*ex*: -57 mV *inh*: -58 mV	**a**
	τ_r	*ex* to *ex* connection: 2 ms *ex* to *inh* connection: 40 ms *inh* to *ex* connection: 80 ms	**a**
	λ	*layer 1*: 0.4 nS *layer 2* to *layer 4*: 1.6 nS	**c**
Learning (Eq 8 -10)	ρ	50.0	*
	τ_C	5 ms	**b**
	τ_D	5 ms	**b**
	α_C	0.5	**b**
	α_D	0.5	**b**

Sources of values adopted: **a** from Troyer et al., 1998; **b** from Perrinet et al., 2001; **c** from Eguchi et al., 2018; * are tuned parameter for simulations; *ex*: excitatory neuron; *inh*: inhibitory neuron.

Appendix II: Detailed Code of the In Silico Visual Cortex

(Omitted)