ARTICLE **Communicated by Tahereh Toosi**

Simulated Complex Cells Contribute to Object Recognition Through Representational Untangling

Mitchell B. Slapik
mslapik@gmail.com
Department of Neurobiology and Anatomy, McGovern Medical School, University of Texas–Houston, Houston, TX 77030, USA

Harel Z. Shouval
harel.shouval@uth.tmc.edu
Department of Neurobiology and Anatomy, McGovern Medical School, University of Texas–Houston, Houston, TX 77030, USA; and Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005, USA

The visual system performs a remarkable feat: it takes complex retinal activation patterns and decodes them for object recognition. This operation, termed "representational untangling," organizes neural representations by clustering similar objects together while separating different categories of objects. While representational untangling is usually associated with higher-order visual areas like the inferior temporal cortex, it remains unclear how the early visual system contributes to this process—whether through highly selective neurons or high-dimensional population codes. This article investigates how a computational model of early vision contributes to representational untangling. Using a computational visual hierarchy and two different data sets consisting of numerals and objects, we demonstrate that simulated complex cells significantly contribute to representational untangling for object recognition. Our findings challenge prior theories by showing that untangling does not depend on skewed, sparse, or high-dimensional representations. Instead, simulated complex cells reformat visual information into a low-dimensional, yet more separable, neural code, striking a balance between representational untangling and computational efficiency.

1 Introduction

The visual system processes complex activation patterns on the retina and categorizes them for object recognition. This feat presents an apparent paradox. On the one hand, the visual system must be exquisitely sensitive

Mitchell Slapik is the corresponding author.

Neural Computation **38, 145–164** (2026)
https://doi.org/10.1162/NECO.a.1480

to fine details of visual information, such as the tiny brushstrokes of a painting or the lines in a friend's expression. On the other hand, it must be robust or "invariant" to low-level changes, such as lighting, scaling, translation, rotation, and pose, which change an object's appearance but not its identity (DiCarlo & Cox, 2007). It is challenging to simultaneously achieve both selectivity and invariance.

Object recognition involves a hierarchy of visual processing stages, progressing from edge detection in primary visual cortex (V1) to shape processing in extrastriate visual cortex (V2 to V4) to object recognition in inferior temporal cortex (IT; Conway, 2018; Felleman & Van Essen, 1991; Lueschow et al., 1994). These transformations are thought to organize neural representations by clustering similar objects together while separating different objects, a process that has been termed "representational untangling" (see Figure 1A). Several strategies for representational untangling have been proposed, including skewed, sparse, and high-dimensional representations. Biologically plausible learning rules that maximize skewness or sparseness generate edge detectors that look remarkably like V1 cells (Albesa-González et al., 2022; Blais et al., 1998; Olshausen & Field, 1997), suggesting that the visual system might rely on skewness or sparseness to solve object recognition. Conversely, it has been argued that the brain aims to add dimensions to our neural representation (Bernardi et al., 2020; Fusi et al., 2016). Such high-dimensional representations allow for many possible linear decision boundaries, enabling downstream neurons to easily read out many different kinds of information. This makes high-dimensional representations useful not just for one task, but for any potential task we may have to perform.

In this study, we investigate how a computational model of early vision contributes to object recognition. We test if it makes stimuli more accessible to linear classification and if it uses skewed, sparse, or high-dimensional representations to accomplish this. To assess the role of the early visual system in representational untangling, we use a simplified computational model of the retina, lateral geniculate nucleus (LGN), simple cells and complex cells (see Figure 1B) and show these models two standard data sets of images, consisting of numerals and objects, respectively (Adelson & Bergen, 1985; Feng et al., 2007; LeCun & Cortes, 2010; Li et al., 2023; Riesenhuber & Poggio, 1999). We find that the simulated complex cells contribute to the untangling of object classes, making linear classification easier on the training data and more robust on the test data. However, this operation does not rely on any previously proposed mechanism, such as skewed, sparse, or high-dimensional representations. Instead, it condenses visual information to a lower-dimensional code while also making it more accessible to classification.

Our study follows previous research that demonstrates the importance of the early visual system for representational untangling. Our work

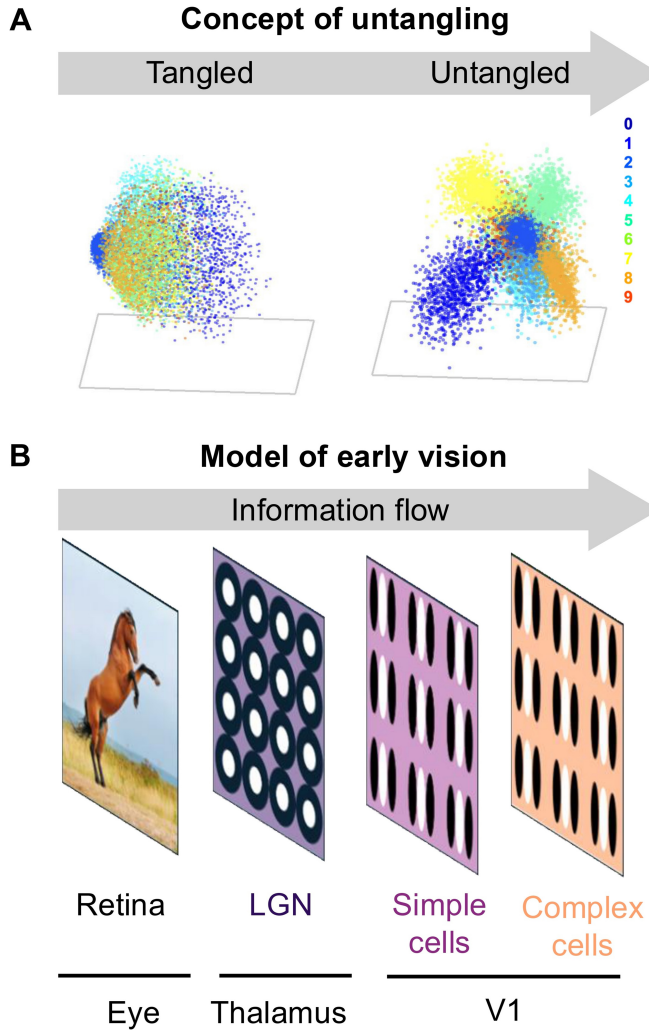

Figure 1: Model of the early visual system and concept of untangling. (A) Demonstrating the concept of untangling using a representation of handwritten digits (0–9) in 3D space. In a tangled space (left), the representations of the different digits overlap. In the untangled space (right), the different digits are represented by minimally overlapping clusters. Here each different color represents a different class. (B) Our model consists of four stages: the retina, lateral geniculate nucleus (LGN), simple cells, and complex cells. Anatomically, this pathway runs from the eyes, through thalamus, into the back of the brain at V1. Here, black represents retina, dark purple represents LGN, light purple represents simple cells, and orange represents complex cells.

builds on these studies by including multiple levels of the visual hierarchy and investigating the mechanism behind this representational untangling (Bergstra et al., 2011; Gáspár et al., 2019; Shams & von der Malsburg, 2002). Ultimately, we believe our results are complementary to these studies, providing further evidence that the early visual system is optimized for representational untangling rather than maximizing information.

2 Methods

2.1 Model of Early Visual Processing. We simulate the early visual system progressing from photoreceptors to the LGN to simple cells and finally to complex cells (see Figure 1B). Retinal photoreceptors project through retinal ganglion cells to the LGN, which forms circular center-surround receptive fields with either an excitatory center and an inhibitory surround (ON cell) or an inhibitory center and excitatory surround (OFF cell) (Kuffler, 1953). LGN then projects to simple cells in V1, which combine several of these circles together into a line in a particular location and orientation (Bonin et al., 2005; Jeffries et al., 2014; Lian et al., 2021; Mechler & Ringach, 2002). Finally, complex cells pool together multiple simple cells of the same orientation but at different locations, responding to a correctly oriented line regardless of its exact location within the receptive field (see Figure 2A; Hubel & Wiesel, 1962).

All of these cell types have standardized computational models. Retinal photoreceptors are modeled by passing the pixel value through a sigmoid nonlinearity. We call this the "retinal representation." Meanwhile, retinal ganglion cells in the LGN can be modeled by the difference of two gaussian curves, one slightly wider than the other (De Valois et al., 2000; Gabbiani & Cox, 2010). Here, we used a standard deviation of 1 pixel for the inner gaussian ($\sigma_{x1}^2,\ \sigma_{y1}^2 = 1$ pixel) and 2 pixels for the outer gaussian ($\sigma_{x2}^2,\ \sigma_{y2}^2 =$ 2 pixels). We call this representation the LGN:

$$(x, y) = \frac{1}{2\pi\sigma_{x1}\sigma_{y1}} \exp\left(-\frac{x^2}{2\sigma_{x1}^2} - \frac{y^2}{2\sigma_{y1}^2}\right) - \frac{1}{2\pi\sigma_{x2}\sigma_{y2}} \exp\left(-\frac{x^2}{2\sigma_{x2}^2} - \frac{y^2}{2\sigma_{y2}^2}\right). \tag{2.1}$$

Simple cells can be modeled by a two-dimensional normal curve multiplied by a sinusoidal wave, also known as a Gabor filter (Gabbiani & Cox, 2010; Jones & Palmer, 1987). In this study, simple cells were simulated at eight different orientations, ranging from $\theta = 0$ radians (horizontal) to $\theta = \frac{7\pi}{8}$ radians at $\frac{\pi}{8}$ radian intervals. We used a standard deviation of 3 pixels ($\sigma_x^2,\ \sigma_y^2 = 3$ pixels), a spatial frequency of 0.8 ($k = 0.8$), and no offset

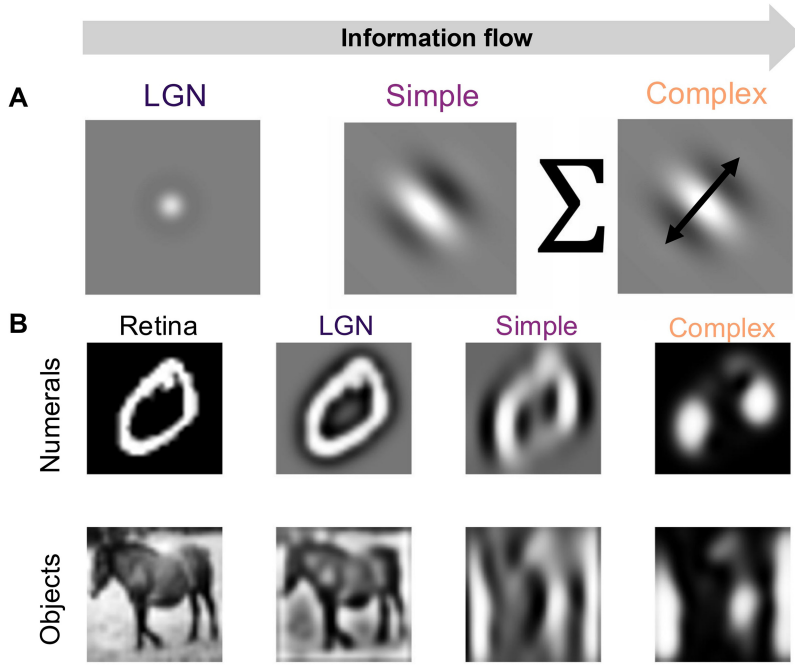

Figure 2: Receptive fields at each stage of the visual hierarchy along with corresponding neural activations. (A) The visual hierarchy progresses from center-surround receptive fields (LGN) to edge detectors (V1). Within V1, simple cells respond to a line at a particular location in the receptive field, whereas complex cells respond to that line at any location within the receptive field. (B) Examples from the numerals (top, 0) and objects (bottom, horse) data sets, along with their corresponding activation maps throughout the visual hierarchy. Here, LGN responds to bright circles surrounded by a dark ring (ON cells), while simple and complex cells respond to vertical edges in the image. The simple and complex cell images are for a single orientation. Our model offers eight orientations. Black represents retina, dark purple represents LGN, light purple represents simple cells, and orange represents complex cells.

($\phi_x = 0$ radians).

$$\omega(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right) \cos\left(\mathrm{k}\left(x\cos(\theta) + y\sin(\theta) - \phi\right)\right). \quad (2.2)$$

Finally, complex cells (R_{cc}) are modeled by combining two simple cells with a phase offset of $\phi = \frac{\pi}{2}$ radian and then squared and summed

(Gabbiani & Cox, 2010; Touryan et al., 2005). These are referred to as "simple cell: even" (R_{se}, $\phi = 0$ radians) and "simple cell: odd" (R_{so}, $\phi = \frac{\pi}{2}$ radians). Again, complex cells were simulated at eight different orientations, ranging from $\theta = 0$ radians (horizontal) to $\theta = \frac{7\pi}{8}$ radians at $\frac{\pi}{8}$ radian intervals.

$$R_{cc} = R_{se}^2 + R_{so}^2. \tag{2.3}$$

Then these neural populations—retina, LGN, simple and complex cells—were all z-scored (Z) and passed through a sigmoid nonlinearity to obtain the final activation value (A). For these steps, we used the following equations, where μ is the mean response for each neuron and σ is the standard deviation:

$$Z = \frac{x - \mu}{\sigma}, \tag{2.4}$$

$$A = \frac{1}{1 + e^{-Z}}. \tag{2.5}$$

We exposed this model to two sets of stimuli: MNIST, which consists of hand-drawn digits 0 to 9, and CIFAR-10, which consists of images of everyday objects, including planes, cars, and birds (Feng et al., 2007; LeCun & Cortes, 2010) (see Figure 2B). For each data set, we simulated a cell centered on every pixel of the input image. The MNIST stimuli, which are 28 by 28 pixels, resulted in 784 simulated LGN cells (28 pixels × 28 pixels) and 6272 simulated simple and complex cells (28 pixels × 28 pixels × 8 angles). The CIFAR-10 stimuli, which are 32 by 32 pixels, resulted in 1024 simulated LGN cells (32 pixels × 32 pixels) and 8192 simulated simple and complex cells (32 pixels × 32 pixels × 8 angles). To obtain the activation values, we multiplied the receptive field of each neuron by the pixel values at each location, z-scored them, and then passed them through a sigmoid nonlinearity.

2.2 Linear decoders. To quantify the untangling of different stimuli, we used a simple linear decoder: linear discriminant analysis (LDA; Fisher, 1936). We trained a linear decoder on 10,000 examples and tested on 10,000 examples, repeating this analysis 30 times for different training and testing data sets.

2.3 Statistics. Skewness and kurtosis were computed separately for each neuron and then averaged across the entire population at each stage of the visual hierarchy. We computed skewness with the following equation:

$$\text{Skewness} = \frac{E(x - \mu)^3}{\sigma^3}. \tag{2.6}$$

And we computed kurtosis as follows:

$$\text{Kurtosis} = \frac{E(x - \mu)^4}{\sigma^4}. \tag{2.7}$$

Finally, we measured separation between stimuli using the Fisher discriminant ratio (FDR), defined as the variance between classes ($\sigma_{between}$) divided by the variance within classes (σ_{within}) (Chen, 2020):

$$FDR = \frac{\sigma^2_{between}}{\sigma^2_{within}}. \tag{2.8}$$

To examine how the early visual cortex achieves representational untangling, linear decoder weights were averaged for each neuron across all possible binary decisions, $\frac{9 \times 10}{2!} = 45$, after taking the absolute value. Then a Pearson correlation was computed between each neuron's average weight and its skewness, kurtosis, or FDR.

We estimated the dimensionality of neural representations using the participation ratio (Gao et al., 2017), where λ represents the eigenvalues:

$$\text{Participation ratio} = \frac{(\sum \lambda)^2}{\sum \lambda^2}. \tag{2.9}$$

We also defined a novel measure called "coding dimensionality," which describes the number of linear dimensions needed to separate object classes rather than measuring the variation within classes. To compute this measure, we applied an equation similar to the participation ratio but based on the Fisher discriminant ratios (FDR) instead of the eigenvalues. We calculate coding dimensionality as follows:

$$\textit{Coding dimensionality} = \frac{(\sum FDR)^2}{\sum FDR^2}. \tag{2.10}$$

3 Results

In this study, we investigated object recognition using a simulated model of the early visual cortex and two data sets: numerals (MNIST) and objects (CIFAR-10). Briefly, we used linear classifiers to decode stimuli at each level of the visual hierarchy. Then we measured various characteristics of the neural representations, including skewness, kurtosis, and dimensionality, to better understand the mechanism behind object recognition. Finally, we correlated these measures with linear decoder weights to see which features better separate different object categories.

3.1 Simulated Complex Cells Perform Representational Untangling of Object Categories. We hypothesized that the complex cells make

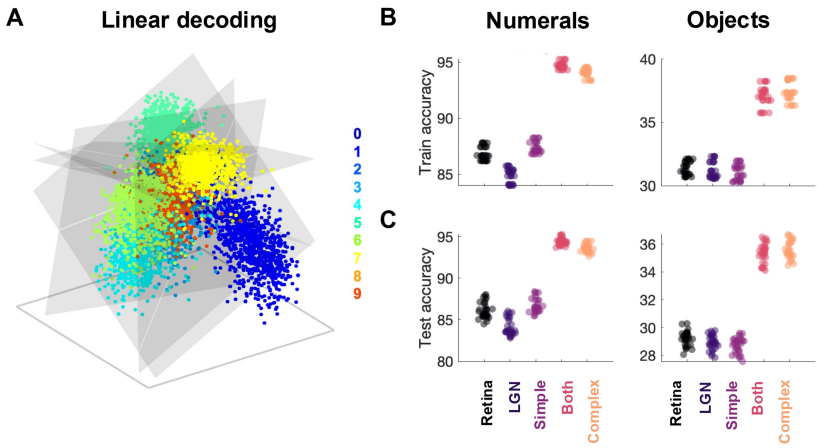

Figure 3: Simulated complex cells improve linear decoding of stimulus identity. (A) LDA visualization of complex cell neural activations for different hand-drawn digits and linear decision boundaries. Here, each circle represents a stimulus, and the color represents a class. The gray planes depict linear decision boundaries. (B) Training set decoding accuracy. The linear decoding accuracy in the four different levels of the early visual system hierarchy. Complex cells have higher training accuracy than previous stages of the simulated visual hierarchy on numerals (left) and objects (right). (C) Test set decoding accuracy, by level. Complex cells have higher testing accuracy than previous stages of the simulated visual hierarchy on numerals (left) and objects (right). Here, each circle represents a model. As before, black represents retina, dark purple represents LGN, light purple represents simple cells, and orange represents complex cells.

different classes more linearly separable than they are on the retina, performing the first steps of object recognition. We tested this hypothesis by linearly decoding stimulus identity based on the simulated activations of retina, LGN, simple cells, and complex cells (see Figure 3). We also included a combined population of simple and complex cells to see if this resulted in a higher accuracy than either one alone. Our data showed that complex cells perform best at separating both drawn digits and natural images in the training and testing data sets (see Figures 3B and 3C). Specifically, in the numerals data set, complex cells had significantly higher training accuracies than the retina (one-way ANOVA, $p = 9.96 \times 10^{-131}$, post hoc two-sample t-test with Bonferroni correction, $p = 1.32 \times 10^{-51}$), LGN ($p = 9.45 \times 10^{-56}$), and simple cells ($p = 1.71 \times 10^{-52}$), but lower accuracy than a combined population of both simple and complex cells ($p = 1.49 \times 10^{-10}$). Likewise, in the objects data set, complex cells had significantly higher accuracy than the retina (one-way ANOVA, $p = 1.29 \times 10^{-94}$, post

hoc two-sample t-test with Bonferroni correction, $p = 1.00 \times 10^{-40}$), LGN ($p = 2.71 \times 10^{-39}$), and simple cells ($p = 1.51 \times 10^{-40}$), but no significant difference from a population of both simple and complex cells ($p = 0.35$), indicating that they had similar accuracies.

Here, complex cells may have had higher accuracy because they were overfitting the training data set. To test this, we examined their accuracy on the held-out test set but found similar results. For the numerals data set, complex cells had significantly higher test accuracy than the retina (one-way ANOVA, $p = 4.09 \times 10^{-106}$, post hoc two-sample t-test with Bonferroni correction, $p = 6.38 \times 10^{-42}$), LGN ($p = 9.95 \times 10^{-47}$), and simple cells ($p = 2.68 \times 10^{-42}$), but significantly lower test accuracy than a combined population of both simple and complex cells ($p = 3.22 \times 10^{-8}$). Meanwhile, in the objects data set, complex cells had significantly higher test accuracy than the retina (one-way ANOVA, $p = 3.13 \times 10^{-107}$, post hoc two-sample t-test with Bonferroni correction, $p = 1.66 \times 10^{-44}$), LGN ($p = 1.23 \times 10^{-45}$), and simple cells ($p = 4.64 \times 10^{-46}$), but did not differ from a population of both simple and complex cells ($p = 0.67$), indicating that they had similar test accuracy. See supplemental Figure 1 for results with different forms of normalization, supplemental Figure 2 for results with different nonlinearities, and supplemental Figure 3 for results with different linear classifiers.

Overall, simulated complex cells showed superior performance in separating both drawn digits and natural images in both training and testing data, suggesting that the early visual cortex performs the first stages of untangling the manifold of natural images. This untangling process mirrors the kernel trick in machine learning, which adds nonlinear dimensions to make classes more linearly separable than they are in the input data. Notably, training and testing accuracy does not steadily increase as we move up the visual hierarchy. Instead, we observe a plateau followed by a sudden improvement with complex cells. We also find differences between the two data sets. Linear decoders achieved significantly higher accuracy for numerals than objects, highlighting the relative ease of decoding numerals over natural objects. In addition, in the numerals data set, a combined population of simple and complex cells performs better than complex cells alone, indicating that the exact position of an edge can be used to decode the hand-drawn digit. This is not the case for natural images, where a combined population of simulated simple and complex cells has no additional advantage over complex cells alone. This indicates that the precise location of an edge did not improve stimulus decoding in the objects data set; instead, the same object can appear in different locations in each image, better mimicking natural stimuli.

3.2 Simulated Complex Cells Increase the Skewness and Sparseness of Neural Representations. We have shown that a model of the early visual system performs "representational untangling" by grouping similar stimuli together and separating different stimuli. To understand the

mechanism behind this representational untangling, we examined two measures of stimulus selectivity across the visual hierarchy: skewness and kurtosis. Skewness is known to result from biologically plausible learning rules such as the Bienenstock-Cooper-Munro theory (BCM; Blais et al., 1998). Meanwhile, sparse filtering on natural images generates edge-detectors that look remarkably like V1 simple cells (Bell & Sejnowski, 1995, 1997; Olshausen & Field, 1997). Sparse distributions mean that neurons have low activations for most stimuli but unusually high values for a small subset of stimuli. This results in a heavily "tailed" distribution in the neural representation, which can be measured by the kurtosis of the distribution.

To examine if skewness or kurtosis contributes to representational untangling, we measured both at each stage of the simulated visual hierarchy. We find that complex cells have greater skewness and kurtosis than previous levels of the hierarchy, providing a potential explanation for how they better separate object categories (see Figures 4C and 4D). Specifically, in the numerals data set, complex cells had significantly higher skewness than LGN (one-way ANOVA, $p = 3.84 \times 10^{-125}$, post hoc two-sample t-test with Bonferroni correction, $p = 5.66 \times 10^{-62}$) and simple cells ($p = 6.00 \times 10^{-80}$), but lower skewness than the retina ($p = 3.82 \times 10^{-52}$). Meanwhile, in the objects data set, complex cells had significantly higher skewness than the retina (one-way ANOVA, $p = 4.51 \times 10^{-187}$, post hoc two-sample t-test with Bonferroni correction, $p = 4.36 \times 10^{-92}$), LGN cells ($p = 1.49 \times 10^{-100}$), and simple cells ($p = 6.88 \times 10^{-98}$). Similarly, we found that complex cells have significantly higher kurtosis than previous stages of the visual hierarchy. In the numerals data set, complex cells had significantly higher kurtosis than LGN (one-way ANOVA, $p = 1.33 \times 10^{-100}$, post hoc two-sample t-test with Bonferroni correction, $p = 1.92 \times 10^{-20}$) and simple cells ($p = 2.14 \times 10^{-43}$), but lower kurtosis than the retina (3.21×10^{-46}). Similarly, in the objects data set, complex cells had significantly higher kurtosis than the retina (one-way ANOVA, $p = 4.10 \times 10^{-169}$, post hoc two-sample t-test with Bonferroni correction, $p = 9.21 \times 10^{-87}$), LGN cells ($p = 6.12 \times 10^{-79}$), and simple cells ($p = 2.49 \times 10^{-80}$). See supplemental Figure 4 for results with different forms of normalization and supplemental Figure 5 for results with different nonlinearities.

Here we notice an important difference between the two data sets, numerals and objects. Numerals consist of black and white pixels, resulting in high skewness and kurtosis in their retinal activations, exceeding even complex cells. Meanwhile, objects have lower skewness and kurtosis in their retinal activations, better reflecting the statistics of natural images. Nevertheless, with the exception of retina, we still find consistent results between the two data sets, where complex cells have greater skewness and kurtosis than LGN and simple cells. This provides a potential explanation for how complex cells better separate object categories.

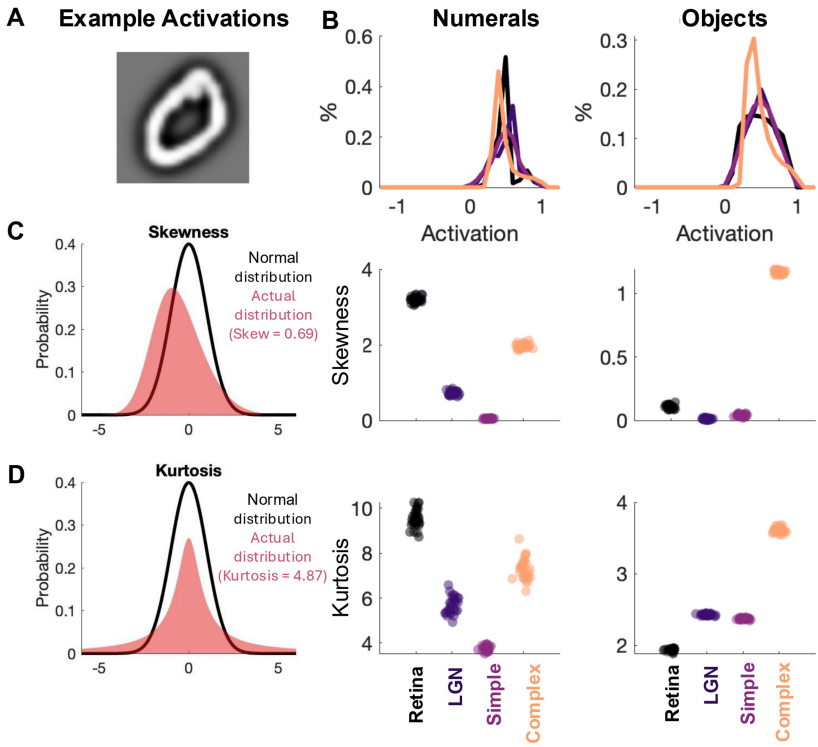

Figure 4: Simulated complex cells increase skewness and sparseness. (A) Example LGN activations for a hand-drawn "0" from the numerals data set. (B) Distribution of activations at different stages of the simulated visual hierarchy for numerals (left) and objects (right). (C) Left: Example of a skewed distribution (red) compared with a normal distribution (black). Complex cells have greater skewness than other stages of the simulated visual hierarchy for numerals (center) and objects (right). (D) Left: Example of a distribution with high kurtosis (red) compared with a normal distribution (black). Complex cells have greater kurtosis than other stages of the simulated visual hierarchy for numerals (center) and objects (right). Here, black represents retina, dark purple represents LGN, light purple represents simple cells, and orange represents complex cells. Each dot represents the average skewness or kurtosis across all cells in each model.

4.3 Skewness and Sparseness Do Not Contribute to Representational Untangling of Object Categories. To test if skewness and kurtosis contribute to representational untangling, we examined if linear decoders tend to emphasize neurons with high skewness, sparseness, or FDR, resulting in higher weights on those neurons. We correlated each neuron's skewness,

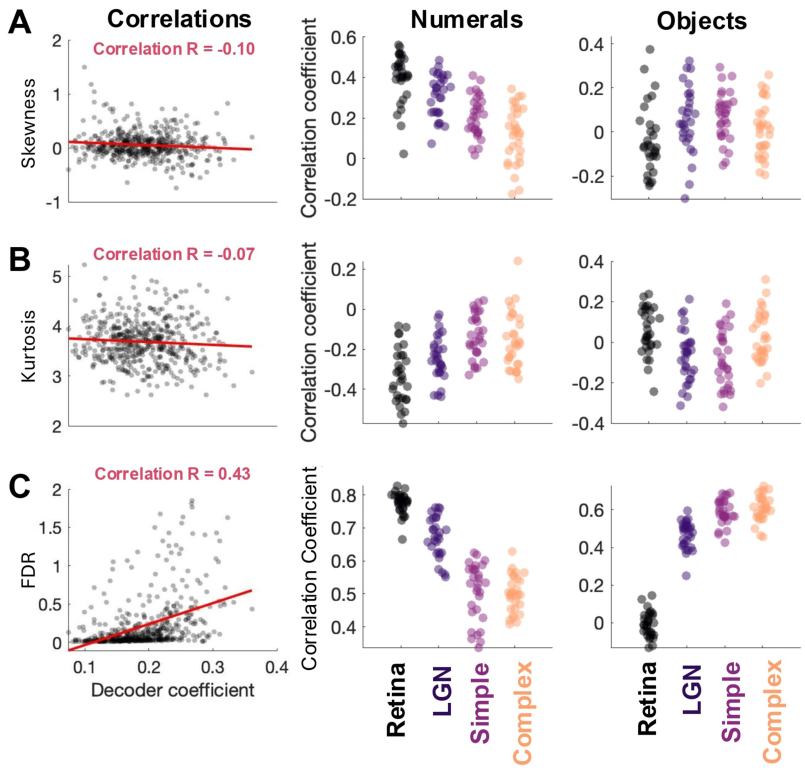

Figure 5: Skewness and sparseness did not contribute to linear decoding of stimulus identity. (A) Left: Example correlation between decoder weights and activation skewness for complex cells on numerals. Here, each dot represents an input neuron. Skewness had weakly positive correlations with linear decoder weights for numerals (center) but nonsignificant correlations for objects (right). Here, each dot represents a model. (B) Left: Example correlation between decoder weights and activation kurtosis for complex cells on numerals. Here, each dot represents a neuron. Kurtosis had nonsignificant or negative correlations with linear decoder weights for numerals (center) and objects (right). (C) Left: Example correlation between decoder weights and FDR for complex cells on numerals. Here, each dot represents a neuron. FDR had strong positive correlations with linear decoder weights for numerals (center) and objects (right). Here, each dot represents a model. Black represents retina, dark purple represents LGN, light purple represents simple cells, and orange represents complex cells.

sparseness, and FDR with the magnitude of their corresponding linear decoder weight (see Figure 5, left column). Here, we generally found weak or nonsignificant correlations between skewness/sparseness and decoder weights, suggesting that the brain does not use skewed or sparse

representations to solve object recognition (see Figures 5A and 5B). In particular, for the numerals data set, we found weakly positive correlations between decoder weights and skewness for the retina ($R = 0.40$, $p = 1.33 \times 10^{-16}$), LGN ($R = 0.31$, $p = 1.14 \times 10^{-15}$), simple cells ($R = 0.21$, $p = 2.38 \times 10^{-11}$) and complex cells ($R = 0.12$, $p = 4.28 \times 10^{-4}$). Meanwhile, for the objects data set, all correlations between decoder weights and skewness were weakly positive for simple cells ($R = 0.071$, $p = 5.47 \times 10^{-3}$) but nonsignificant for retina ($R = -0.035$, $p = 0.88$), LGN ($R = 0.068$, $p = 0.12$) and complex cells ($R = 0.010$, $p = 1$).

We believe that these weak or nonsignificant correlations, especially on the objects data set, indicate that skewness does not play a central role in representational untangling. Similarly, correlations between decoder weights and kurtosis were nonsignificant or negative. For the numerals data set, correlations between decoder weights and kurtosis were significantly negative for the retina ($R = -0.33$, $p = 2.78 \times 10^{-13}$), LGN ($R = -0.24$, $p = 2.37 \times 10^{-12}$), simple cells ($R = -0.13$, $p = 1.07 \times 10^{-6}$), and complex cells ($R = -0.16$, $p = 3.71 \times 10^{-6}$). Meanwhile, in the objects data set, correlations between decoder weights and kurtosis were nonsignificant for retina ($R = 0.052$, $p = 0.073$) and complex cells ($R = 0.026$, $p = 1$) but significantly negative for LGN ($R = -0.067$, $p = 0.032$) and simple cells ($R = -0.082$, $p = 0.012$). Overall, we generally found weak or nonsignificant correlations between decoder weights and skewness/kurtosis, indicating that the brain does not use skewed or sparse representations to solve object recognition. If anything, linear decoders tended to use neurons with low kurtosis.

Finally, we correlated decoder weights with neurons' Fisher discriminant ratio (FDR), which measures separation between different classes (Chen, 2020). This analysis yielded high correlations in both data sets (see Figure 5C). Specifically, in the numerals data set, we found significantly positive correlations for the retina ($R = 0.77$, $p = 4.76 \times 10^{-41}$), LGN ($R = 0.67$, $p = 4.47 \times 10^{-31}$), simple cells ($R = 0.51$, $p = 1.55 \times 10^{-23}$), and complex cells ($R = 0.50$, $p = 5.18 \times 10^{-29}$). Likewise, for the objects data set, correlations with FDR were significantly positive for LGN ($R = 0.47$, $p = 6.78 \times 10^{-25}$), simple cells ($R = 0.59$, $p = 1.03 \times 10^{-28}$), and complex cells ($R = 0.61$, $p = 9.87 \times 10^{-29}$) but nonsignificant for retina ($R = -0.0078$, $p = 1$). Therefore, rather than emphasizing skewness or kurtosis, we found that linear decoders predominantly emphasized neurons with high stimulus separation as measured by FDR. See supplemental Figure 6 for results with different forms of normalization, supplemental Figure 7 for results with different nonlinearities, and supplemental Figure 8 for correlations with max weights. Finally, see supplemental Figure 8 for a causal analysis that compares decoding accuracy between high and low skewness/kurtosi/FDR subpopulations.

4.4 Simulated Complex Cells Condense Information into a Low-Dimensional Representation. Finally, we aimed to test the hypothesis that the brain uses high-dimensional representations to better untangle natural

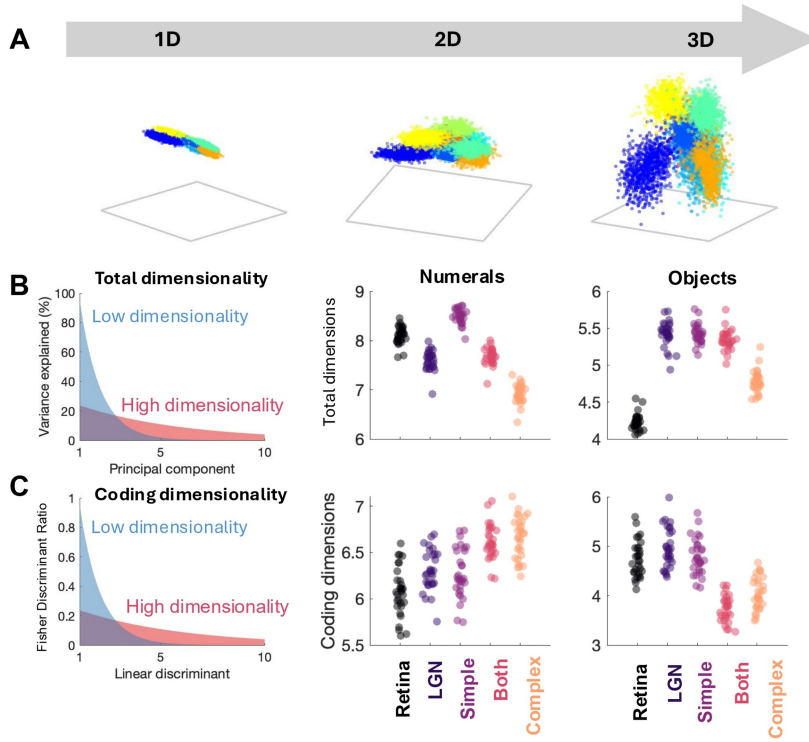

Figure 6: Simulated complex cells reduce total dimensionality of neural representations. (A) Schematic depicting increasing dimensionality of a neural representation. (B) Left: Schematic depicting eigenspectra for a low-dimensional and high-dimensional neural representations. Complex cells reduce total dimensionality for numerals (center) and objects (right). (C). Left: Schematic depicting linear discriminants for low-dimensional and high-dimensional neural representations. Complex cells increase coding dimensionality for numerals (center) but reduce it for objects (right), reflecting a difference between hand-drawn digits and natural images. Here, black represents retina, dark purple represents LGN, light purple represents simple cells and orange represents complex cells.

images (see Figure 6A). Prior work has shown that high-dimensional representations permit more possible decision boundaries, making them more adept than low-dimensional representations at separating different stimuli (Bernardi et al., 2020; Fusi et al., 2016).

To test this idea, we examined the dimensionality of neural representations at different stages of the simulated visual hierarchy. We used the participation ratio method (see equation 2.9) to approximate the linear dimensionality of representations in the retina, LGN, simple cells, and complex

cells. We also included a combined population of simple cells and complex cells. We found a steady decrease in the dimensionality of the neural code as we move up the simulated visual hierarchy (see Figure 6B). In particular, for the numerals data set, complex cells had significantly lower dimensionality than the retina (one-way ANOVA, $p = 5.02 \times 10^{-72}$, post hoc two-sample t-test with Bonferroni correction, $p = 3.84 \times 10^{-32}$), LGN cells ($p = 6.65 \times 10^{-19}$), simple cells ($p = 2.30 \times 10^{-40}$), and a combined population of both simple and complex cells ($p = 6.96 \times 10^{-23}$). Similarly, for the objects data set, complex cells had significantly lower dimensionality than LGN (one-way ANOVA, $p = 1.65 \times 10^{-78}$, post hoc two-sample t-test with Bonferroni correction, $p = 1.90 \times 10^{-21}$), simple cells ($p = 1.84 \times 10^{-24}$), and a combined population of both simple and complex cells ($p = 3.07 \times 10^{-21}$), but higher dimensionality than retina ($p = 1.02 \times 10^{-22}$). The objects data set may have an artificially low dimensionality in the retina due to the effect of brightness, which dominates the first principal component of natural images.

Meanwhile, we also defined a novel measure, coding dimensionality (see equation 2.10), which describes the number of dimensions needed to separate object classes. Here, we find mixed results, with complex cells increasing coding dimensionality for numerals but decreasing it for objects (see Figure 6C). Specifically, for numerals, complex cells had significantly higher coding dimensionality than the retina (one-way ANOVA, $p = 6.96 \times 10^{-20}$, post hoc two-sample t-test with Bonferroni correction, $p = 1.51 \times 10^{-12}$), LGN ($p = 1.92 \times 10^{-7}$), and simple cells ($p = 3.85 \times 10^{-7}$), but no significant difference from a combined population of both simple and complex cells ($p = 1$). However, for the objects data set, complex cells showed lower dimensionality than the retina (one-way ANOVA, $p = 1.68 \times 10^{-33}$, post hoc t-test with Bonferroni correction, $p = 3.50 \times 10^{-10}$), LGN ($p = 2.40 \times 10^{-13}$), simple cells ($p = 5.23 \times 10^{-11}$), but higher dimensionality than a population of both simple cells and complex cells ($p = 6.07 \times 10^{-4}$). Although unexpected, we confirmed these findings with several supplemental analyses (not shown). This discrepancy between the two data sets in the complex cell dimensionality likely reflects the difference between written characters and natural images. See supplemental Figure 10 for results with different forms of normalization and supplemental Figure 11 for results with different nonlinearities.

4 Discussion

Our findings provide evidence that simulated complex cells play an active role in object recognition, thereby challenging a prevailing view that object recognition is confined to higher visual areas (Conway, 2018; Lueschow et al., 1994). We find that the simulated complex cells significantly contribute to the representational untangling of object classes, making linear classification easier on the training data and more robust on the test data. Although our simulated complex cells continue to have a limited accuracy

in linear classification of natural images (∼35%), this is still significantly better than the accuracy at earlier stages of the visual hierarchy such as the retina, LGN, and simple cells (Gáspár et al., 2019) (∼30%) (see Figure 3C). We can imagine that by repeating this process of nonlinear transformations throughout the visual hierarchy, we could achieve human-level performance on object recognition by the time we reach it (Conway, 2018; DiCarlo & Cox, 2007; Lueschow et al., 1994).

We further explored the mechanisms behind this untangling process, challenging prior hypotheses that it relies on skewed, sparse, or high-dimensional representations (Albesa-González et al., 2022; Bernardi et al., 2020; Blais et al., 1998; Olshausen & Field, 1997). While simulated complex cells displayed increased skewness and sparseness compared to earlier stages of the visual hierarchy, our linear decoder analysis generally revealed no strong preference for these attributes, suggesting that they do not contribute significantly to object recognition.

One might hypothesize that simulated complex cells improve linear decoding by increasing the dimensionality of the neural code, permitting more possible decision boundaries (see Figure 7A; Fusi et al., 2016). Contrary to this hypothesis, our findings show that the simulated visual hierarchy does not systematically add dimensions to our neural representations but instead condenses visual information into a low-dimensional, yet untangled neural code. We find that the final stage of our model, complex cells, has lower dimensionality than any previous stage of the visual hierarchy (see Figure 7B). Crucially, this may provide an inductive bias for the visual system (Goyal & Bengio, 2022). Perhaps the brain does not carry all information forward to higher visual areas. Instead, it only sends more abstract information about orientation content while dropping detailed information about specific locations and pixels. This would mean that the brain does not aim to "shatter dimensionality" (Bernardi et al., 2020) in the sense of distinguishing between every possible pair of stimuli. Instead, it only aims to distinguish between abstract patterns like two different objects rather than detailed patterns like two images of white noise. This possible inductive bias may enable the brain to learn faster and with fewer data. Low-dimensional representations have also been hypothesized to improve generalization and require fewer neural resources (Bernardi et al., 2020; Boyle et al., 2024). Thus, the visual system strikes a balance between representational untangling and computational efficiency, achieving a best-of-both-worlds scenario.

We acknowledge several limitations in this study, including the use of an idealized model, highly curated data sets, and a specific visual task. Although our model is based on biological data, it leaves out many biological variables, including the distinction between excitatory and inhibitory cells, the influence of neurotransmitters and neuromodulators, feedback connections from higher cortical areas, and plasticity over time (Duncan, 2002; Jedlicka, 2002; Juan & Walsh, 2003; King et al., 2013). Exploring these

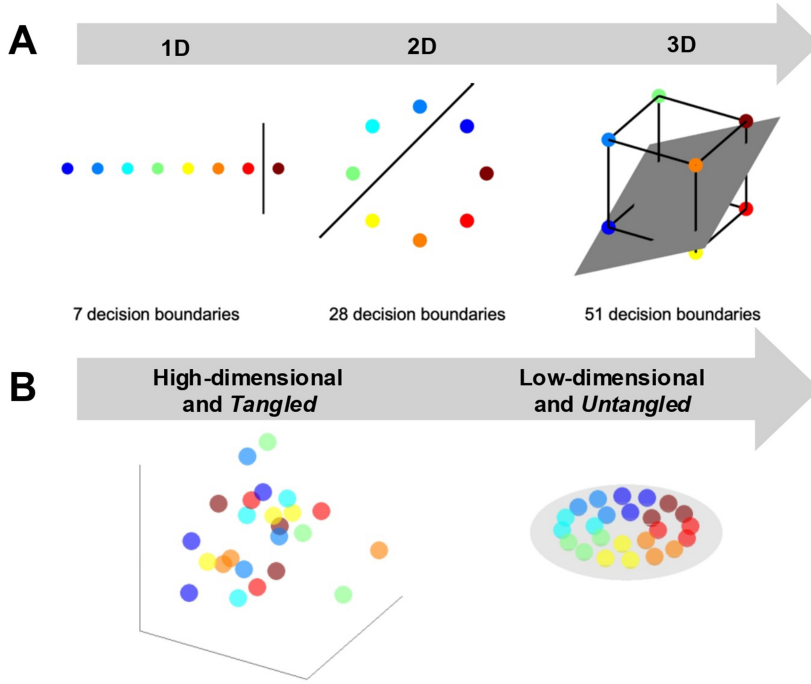

Figure 7: Simulated complex cells simultaneously untangle representations and condense dimensionality. (A) Schematic showing how high-dimensional representations permit more possible decision boundaries, making them useful for a wide variety of tasks. (B) Schematic summarizing our findings: the visual system compresses information into a low-dimensional representation while also untangling it for object recognition. Throughout, each dot represents a stimulus.

elements could reveal additional aspects of representational untangling in the early visual cortex. For example, feedback from higher visual areas or synaptic plasticity may preferentially enhance task-relevant visual features that differentiate object classes. Furthermore, this study used a highly curated data set that lacks the variability of natural scenes such as noise and object occlusion (Gong et al., 2023). Future research could use dynamic videos and multisensory environments to better capture the richness of real-world perception. Third, our image classification task does not explore the diversity of tasks for which vision is used (Zador et al., 2023). For example, we may use visual data to navigate through an environment or read a friend's emotions rather than just identifying semantically relevant objects. Future work can investigate how representational untangling functions outside object recognition, such as navigation and multisensory integration.

Finally, it will be important to examine how representational untangling continues in higher areas like V4 and IT. We hypothesize that representational untangling will steadily continue up the hierarchy until it reaches human behavioral performance in IT. Furthermore, while this study examines how stimulus decoding and computational efficiency are combined in the brain, these findings may equally inform the design of artificial networks, which face similar demands. Specifically, our findings suggest that compressing information in early processing can lead to improved performance. Thus, our findings may reflect a more general principle spanning both artificial and biological networks.

Acknowledgments

The research reported in this article was supported by the National Eye Institute of the National Institutes of Health under award number 5F30EY035603-02 and the UTHealth Houston Center for Clinical and Translational Sciences under award number TL1 TR003169. The content is solely our responsibility and does not necessarily represent the official views of the National Institutes of Health.

References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, *2*(2). 10.1364/josaa.2.000284

Albesa-González, A., Froc, M., Williamson, O., & van Rossum, M. C. W. (2022). Weight dependence in BCM leads to adjustable synaptic competition. *Journal of Computational Neuroscience*, *50*(4). 10.1007/s10827-022-00824-w

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, *7*(6). 10.1162/neco.1995.7.6.1129

Bell, A. J., & Sejnowski, T. J. (1997). The "independent components" of natural scenes are edge filters. *Vision Research*, *37*(23). 10.1016/S0042-6989(97)00121-1

Bergstra, J., Bengio, Y., & Louradour, J. (2011). Suitability of V1 energy models for object classification. *Neural Computation*, *23*(3). 10.1162/NECO_a_00084

Bernardi, S., Benna, M. K., Rigotti, M., Munuera, J., Fusi, S., & Salzman, C. D. (2020). The geometry of abstraction in the hippocampus and prefrontal cortex. *Cell*, *183*(4). 10.1016/j.cell.2020.09.031

Blais, B. S., Intrator, N., Shouval, H., & Cooper, L. N. (1998). Receptive field formation in natural scene environments: Comparison of single cell learning rules. In M. Kearns, S. Solla, & D. Cohn (Eds.), *Advances in neural information processing systems, 11*. MIT Press.

Bonin, V., Mante, V., & Carandini, M. (2005). The suppressive field of neurons in lateral geniculate nucleus. *Journal of Neuroscience*, *25*(47). 10.1523/JNEUROSCI.3562-05.2005

Boyle, L. M., Posani, L., Irfan, S., Siegelbaum, S. A., & Fusi, S. (2024). Tuned geometries of hippocampal representations meet the computational demands of social memory. *Neuron*, *112*(8). 10.1016/j.neuron.2024.01.021

Chen, B. W. (2020). Incomplete data classification—Fisher discriminant ratios versus Welch discriminant ratios. *Future Generation Computer Systems*, *108*. 10.1016/j.future.2018.05.003

Conway, B. R. (2018). The organization and operation of inferior temporal cortex. In *Annual Review of Vision Science*, *4*(1). 10.1146/annurev-vision-091517-034202

De Valois, R. L., Cottaris, N. P., Mahon, L. E., Elfar, S. D., & Wilson, J. A. (2000). Spatial and temporal receptive fields of geniculate and cortical cells and directional selectivity. *Vision Research*, *40*(27). 10.1016/S0042-6989(00)00210-8

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences*, *11*(8). 10.1016/j.tics.2007.06.010

Duncan, J. S. (2002). Neurotransmitters, drugs and brain function. *British Journal of Clinical Pharmacology*, *53*(6). 10.1046/j.1365-2125.2002.01607.x

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1*(1).

Feng, H., Misra, V., & Rubenstein, D. (2007). The CIFAR-10 data set. *Electrical Engineering*, *35*(1).

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics 7*, 179–188. http://dx.doi.org/10.1111/j.1469-1809.1936.tb02137.x.

Fusi, S., Miller, E. K., & Rigotti, M. (2016). Why neurons mix: High dimensionality for higher cognition. *Current Opinion in Neurobiology*, *37*. 10.1016/j.conb.2016.01.010

Gabbiani, F., & Cox, S. J. (2010). *Mathematics for neuroscientists*. Elsevier.

Gao, P., Trautmann, E., Yu, B., Santhanam, G., Ryu, S., Shenoy, K., & Ganguli, S. (2017). *A theory of multineuronal dimensionality, dynamics and measurement*. bioRxiv.

Gáspár, M. E., Polack, P. O., Golshani, P., Lengyel, M., & Orbán, G. (2019). Representational untangling by the firing rate nonlinearity in V1 simple cells. *eLife*, *8*. 10.7554/eLife.43625

Gong, Z., Zhou, M., Dai, Y., Wen, Y., Liu, Y., & Zhen, Z. (2023). A large-scale fMRI data set for the visual processing of naturalistic scenes. *Scientific Data*, *10*(1). 10.1038/s41597-023-02471-x

Goyal, A., & Bengio, Y. (2022). Inductive biases for deep learning of higher-level cognition. In *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *478*(2266). 10.1098/rspa.2021.0068

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

Jedlicka, P. (2002). Synaptic plasticity, metaplasticity and BCM theory. *Bratislavské lekárske listy*, *103*(4–5).

Jeffries, A. M., Killian, N. J., & Pezaris, J. S. (2014). Mapping the primate lateral geniculate nucleus: A review of experiments and methods. *Journal of Physiology Paris*, *108* (1). 10.1016/j.jphysparis.2013.10.001

Jones, J. P., & Palmer, L. A. (1987). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, *58*(6). 10.1152/jn.1987.58.6.1187

Juan, C. H., & Walsh, V. (2003). Feedback to V1: A reverse hierarchy in vision. *Experimental Brain Research*, *150*(2). 10.1007/s00221-003-1478-5

King, P. D., Zylberberg, J., & Deweese, M. R. (2013). Inhibitory interneurons decorrelate excitatory cells to drive sparse code formation in a spiking model of V1. *Journal of Neuroscience*, *33*(13). 10.1523/JNEUROSCI.4188-12.2013

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, *16*(1). 10.1152/jn.1953.16.1.37

LeCun, Y., & Cortes, C. (2010). MNIST handwritten digit database.

AT&T Labs. MNIST database. https://www.kaggle.com/datasets/hojjatk/mnist-dataset, *7*.

Li, Z., Caro, J. O., Rusak, E., Brendel, W., Bethge, M., Anselmi, F., . . . Pitkow, X. (2023). Robust deep learning object recognition models rely on low frequency information in natural images. *PLOS Computational Biology*, *19*(3). 10.1371/journal.pcbi.1010932

Lian, Y., Almasi, A., Grayden, D. B., Kameneva, T., Burkitt, A. N., & Meffin, H. (2021). Learning receptive field properties of complex cells in V1. *PLOS Computational Biology*, *17*(3). 10.1371/journal.pcbi.1007957

Lueschow, A., Miller, E. K., & Desimone, R. (1994). Inferior temporal mechanisms for invariant object recognition. *Cerebral Cortex*, *4*(5). 10.1093/cercor/4.5.523

Mechler, F., & Ringach, D. L. (2002). On the classification of simple and complex cells. *Vision Research*, *42*(8). 10.1016/S0042-6989(02)00025-1

Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, *37*(23). 10.1016/S0042-6989(97)00169-7

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*(11). 10.1038/14819

Shams, L., & von der Malsburg, C. (2002). The role of complex cells in object recognition. *Vision Research*, *42*(22). 10.1016/S0042-6989(02)00202-X

Touryan, J., Felsen, G., & Dan, Y. (2005). Spatial structure of complex cell receptive fields measured with natural images. *Neuron*, *45*(5). 10.1016/j.neuron.2005.01.029

Zador, A., Escola, S., Richards, B., Ölveczky, B., Bengio, Y., Boahen, K., . . . Tsao, D. (2023). Catalyzing next-generation artificial intelligence through NeuroAI. *Nature Communications*, *14*(1). 10.1038/s41467-023-37180-x

Received December 16, 2024; accepted September 19, 2025.