Zhu et al. *Zool. Res.* 2026, 47(3): 960−976
https://doi.org/10.24272/j.issn.2095-8137.2025.581
https://cstr.cn/32248.14.zr.2025.581

ZR | Zoological Research

Article Open Access

Neural representation of object category and viewpoint in the entopallium of pigeons

Jun-Cai Zhu[1], Min-Jie Zhu[1], Peng Wu[1], Qing-Zhi He[1], Jiang-Tao Wang[1,2], Xiao-Ke Niu[1,2], Zhi-Zhong Wang[1,2,*]

[1] *School of Electrical and Information Engineering*, *Zhengzhou University*, *Zhengzhou*, *Henan* 450001, *China*

[2] *Henan Key Laboratory of Brain Science and Brain Computer Interface Technology*, *Zhengzhou*, *Henan* 450001, *China*

ABSTRACT

Object recognition depends on the ability of the visual system to preserve stable category assignment despite variation in viewpoint. Although the mechanisms encoding object category and viewpoint have been extensively described in the primate ventral visual pathway, it remains unclear how the avian brain, which does not contain a layered cortical structure, carries out similarly complex visual computations. Here, we systematically investigated how neurons in the pigeon entopallium (ENTO), the terminal station of the tectofugal visual pathway, represent object identity and viewpoint. Large-scale electrophysiological recordings showed that ENTO neurons displayed moderate category selectivity and strong continuity in viewpoint tuning, and a subset of neurons expressed both tuning properties. At the population level, neural responses formed organized representational manifolds that supported categorical separation and viewpoint continuity at the same time. Moreover, ENTO neurons were strongly sensitive to color, and subpopulation analyses indicated that object representations in the ENTO were not confined to one processing level but jointly included low-level visual features (e.g., color and shape) and higher-level semantic information. These results suggest that the avian visual system can construct complex object representations. During this process, the ENTO may establish functionally specific neural representations through distributed coding based on sparse combinations of distinct neuronal subpopulations.

Keywords: Entopallium; Object recognition; Neural manifold; Electrophysiology; Viewpoint invariance

INTRODUCTION

Object recognition is a central function of the visual system and requires neural mechanisms that preserve stable identity across changes in viewpoint, complex backgrounds, and low-level sensory variability. In primates, extensive studies have shown that the ventral visual pathway gradually abstracts low-level visual features through hierarchical processing and ultimately reaches the inferior temporal cortex (IT), where neurons show category selectivity and viewpoint invariance (Bao et al., 2020; Dicarlo & Cox, 2007; Farhang et al., 2025; Majaj et al., 2015; Nam et al., 2021). By contrast, birds do not possess a laminated neocortex, yet they still show remarkable visual abilities, including discrimination of complex patterns, viewpoint-invariant object recognition, and semantic categorization (Anderson et al., 2020; Levenson et al., 2015; Navarro et al., 2020; Pusch et al., 2022; Soto & Wasserman, 2016). However, whether the avian visual system contains neural representational mechanisms comparable to those found in mammals, and especially how its higher visual areas encode information about object category and viewpoint, remains largely unresolved at the neurophysiological level.

Within the avian visual system, the ENTO is regarded as an important region involved in intermediate- to high-level visual processing. As the terminal stage of the tectofugal pathway, ENTO receives strong thalamic input from the nucleus rotundus (RT) (Benowitz & Karten, 1976; Karten & Hodos, 1970; Straight et al., 2024) and maintains reciprocal connections with the mesopallium ventrolaterale (MVL) (Atoji & Wild, 2012; Krützfeldt & Wild, 2005). It also projects feedforward signals to the nidopallium caudolaterale (NCL), a higher-order associative region that provides top-down feedback to ENTO (Leutgeb et al., 1996; Stacho et al., 2020). Functionally, ENTO neurons possess large receptive fields and respond selectively to low-level visual attributes, including color and motion (Gu et al., 2002; Hsiao et al., 2020; Niu et al., 2024; Xiao et al., 2006), with possible functional specialization among subregions (Clark & Colombo, 2020). However, whether ENTO supports abstract category representations remains controversial (Anderson et al., 2020; Azizi et al., 2019).

To clarify how the ENTO encodes object information, we focused on two basic representational dimensions: object category, which reflects the abstract identity of an object, and viewpoint, which represents continuous geometric transformations across viewing angles. An effective visual

Received: 26 March 2026; Accepted: 15 April 2026; Online: 16 April 2026

Foundation items: This work was supported by the National Natural Science Foundation of China (62206253), and China Postdoctoral Science Foundation (2024M752934)

*Corresponding author, E-mail: wzz1982@zzu.edu.cn

system must preserve category separability while maintaining viewpoint continuity. Using multichannel extracellular recordings combined with quantitative analyses, we examined the following questions: (1) Do ENTO neurons show selective tuning to object category and viewpoint? (2) Are these tuning properties arranged independently or jointly at the single-neuron and population levels? (3) Does the population representation in ENTO support category separability together with viewpoint continuity? (4) How do low- and high-level visual features influence this representational structure?

Our results show that ENTO neurons exhibit robust tuning to both object category and viewpoint, and their population activity forms a high-dimensional representational manifold that combines category separability with viewpoint continuity. Further analyses revealed an association between low-level visual attributes, such as color and shape, and category representations, suggesting that low-level feature information may be combined with higher-level representations within the ENTO of the avian visual system. These findings provide new neurophysiological evidence for representational mechanisms in non-mammalian species and offer cross-species support for a unified geometric framework of visual information encoding.

MATERIALS AND METHODS

Neurophysiological recording

Extracellular recordings were collected from 16 acutely prepared pigeons (Columba livia, both sexes, ~1 year old). After surgical preparation, each pigeon was head-fixed in a custom stereotaxic apparatus (RWD Life Science, China) designed to correspond to the stereotaxic atlas of Karten (Karten & Hodos, 1968). Neural signals were recorded using a 32-channel tungsten microelectrode array (Kedou Brain-Machine Technology, China; 4×8 configuration, uniform 300 μm center-to-center spacing in both row and column directions, 800 kΩ impedance). All recordings were performed in the left ENTO, and the left eye was occluded to restrict visual input to the right hemisphere. Neural activity was amplified, filtered, and digitized with a Cerebus system (Blackrock Microsystems, USA). Because recording sessions were prolonged and isolated single units had limited stability, the analyses were based on multi-unit activity (MUA). All procedures were approved by the Animal Ethics Committee of Zhengzhou University (No. ZZUIRB2022-44) and followed institutional and national guidelines for animal research.

Visual stimuli

The visual stimulus set contained two dimensions: object category and viewpoint (Supplementary Figure S1). The category stimulus set included 200 images from eight object categories (25 images per category): butterfly, elephant, human face, pigeon, backpack, beemug, cowboyhat, and electricguitar. These categories were further classified into animate and inanimate groups. The viewpoint stimulus set included six objects (elephant, human face, pigeon, beemug, cowboyhat, and electricguitar), each rendered from 18 evenly spaced viewpoints covering a 0–360° range. Stimuli were obtained from the Caltech-256 dataset (Griffin et al., 2007) and custom-generated 3D models.

The category stimulus set was deliberately constructed to include categories that are behaviorally or experientially relevant to pigeons (e.g., pigeons, human faces) as well as visually complex but behaviorally neutral categories (e.g., backpack, beer mug, electric guitar). This mixed design allowed evaluation of whether ENTO representations preferentially support category discrimination for ecologically or experientially salient stimuli, while also serving as a conservative control for separating category-related population structure from responses mainly driven by low-level visual features.

All images were resized to a uniform scale, shown against a gray background, and presented using Psychtoolbox (PTB, v.3.0.17) in MATLAB (R2020a). During experiments, the head of each pigeon was fixed about 30 cm from the display (Figure 1A). To accommodate the laterally positioned eyes of pigeons, body orientation and monitor rotation were adjusted to match the natural visual field of pigeons (Erichsen et al., 1989). Before formal stimulus presentation, sparse-noise mapping was used to determine the receptive field (RF) of each neuron, and the screen position was adjusted so that the RF center was aligned with the screen center. The display monitor was color-calibrated, and testing confirmed that the receptive fields of the recorded neurons covered almost the entire screen (about 110° of visual angle). After coarse receptive field localization, test images were further presented at several candidate positions within the receptive field, and the position eliciting the strongest response was selected as the preferred stimulus location for that neuron. Each image subtended about 20° of visual angle on the screen. Visual stimuli were displayed on a 50-inch LCD monitor (AG485UD, AOC, China; resolution: 3 840×2 160 pixels; refresh rate: 100 Hz; pixel density: 36.486 pixels/cm). Stimuli were presented for 500 ms each, with 20 s inter-stimulus intervals to prevent rapid neuronal adaptation (Wang et al., 2024). Each image was repeated three times in randomized order throughout the experiment.

Histological verification

Electrodes were implanted using stereotaxic coordinates AP 10.0 mm, ML 6.0 mm, and DV 4.5 mm, with the anterior portion of the ENTO as the target region. Previous studies have indicated that the anterior ENTO is more closely involved in processing visual features, including shape and color (Clark & Colombo, 2020). To label the recording positions, electrolytic lesions were produced at the end of each recording session (0.3 mA for 30 s). Animals were then euthanized using an overdose of 20% urethane, followed by transcardial perfusion with phosphate-buffered saline (PBS, Biosharp, China) and subsequently 4% paraformaldehyde (PFA, Servicebio, China) for tissue fixation. Brains were removed and post-fixed in 4% paraformaldehyde for 12 h, and then cryoprotected by immersion in 30% sucrose at 4 °C for 48 h. Thereafter, brains were frozen and cut into coronal sections at 20 μm thickness using a cryostat. All sections were subjected to Nissl staining, and electrode tracks and lesion locations were inspected under a light microscope. In previous sessions, all recording positions were verified to be within the ENTO (Supplementary Figure S2). In the present study, five brains were chosen for histological reconstruction, and all recording sites were confirmed to lie within the ENTO (Supplementary Figure S3).

Unit inclusion criteria and quality control

For each stimulus, the mean firing rate of each recorded unit was calculated within a 50–500 ms window after stimulus onset, and baseline activity recorded before stimulus presentation was subtracted. Because rapid repetition of stimuli can induce neural adaptation and reduce response magnitude (Wang et al., 2024), a paradigm with long inter-

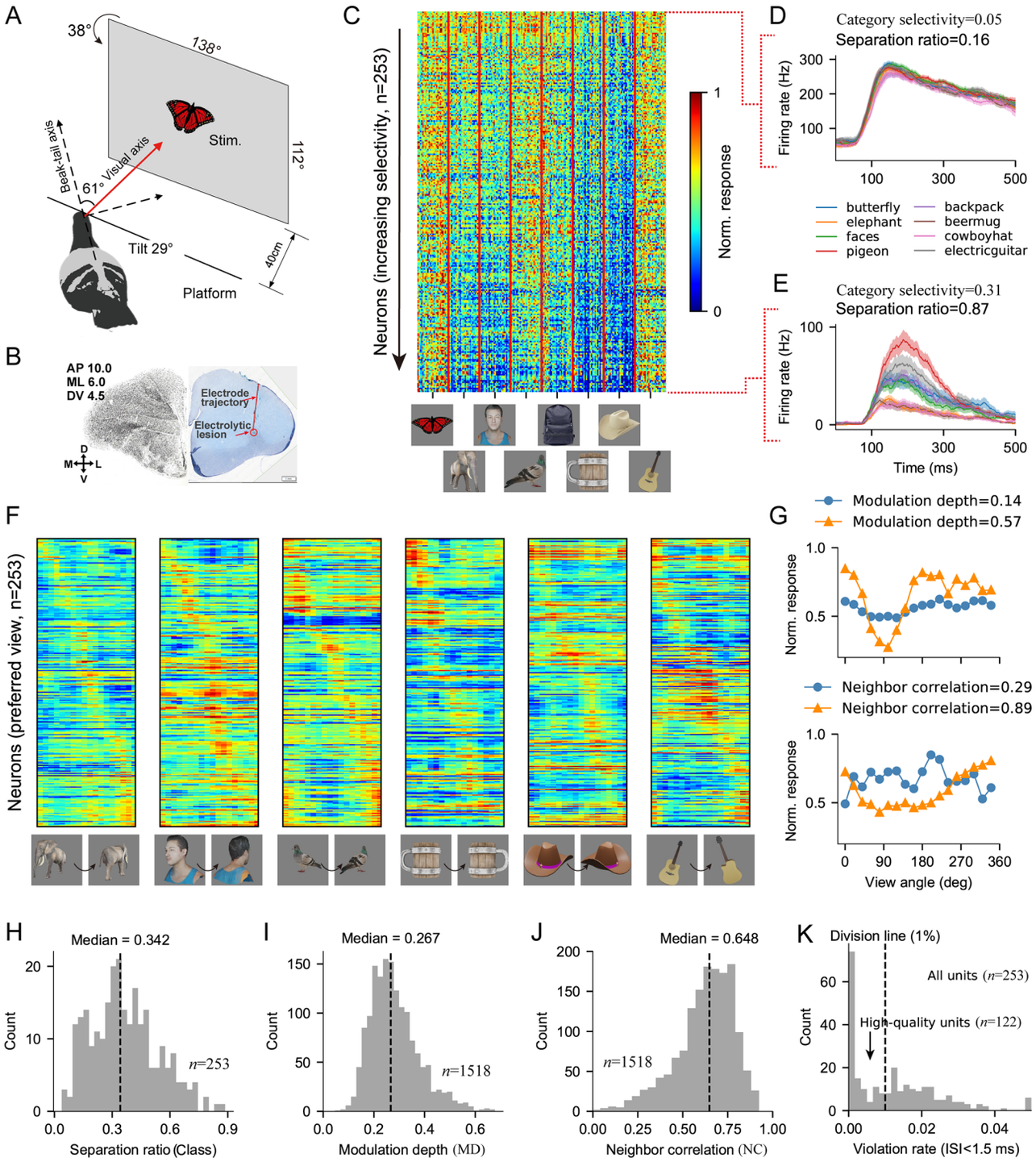

Figure 1 Single-neuron tuning properties in ENTO

A: Schematic illustration of the experimental paradigm. Neural activity was recorded from the ENTO of anesthetized pigeons during passive presentation of visual stimuli. B: Histological reconstruction indicating electrode trajectories and recording locations. C: Normalized responses of 253 reliably responsive neurons to 200 category images (8 categories×25 images), arranged according to increasing category selectivity. D: Representative neuron exhibiting weak category selectivity. E: Representative neuron exhibiting strong category selectivity. F: Normalized responses of the same 253 neurons to 108 viewpoint stimuli (6 objects×18 viewpoints). For each object, neurons were independently reordered based on their preferred viewing angle. G: Examples of viewpoint tuning. Top: neurons with low and high modulation depth (MD=0.14 and 0.57, respectively). Bottom: neurons with low and high tuning continuity, measured by neighbor correlation (NC=0.29 and 0.89, respectively). H: Distribution of category separation ratios across recorded units (*n*=253 units). The dashed line denotes the population median (0.342). I: Distribution of modulation depth for each unit × viewpoint condition (*n*=1518; 253 units across 6 viewpoints). The dashed line marks the population median (0.267). J: Distribution of neighbor correlation. The dashed line marks the population median (0.648). K: Distribution of refractory-period violation rates (ISI<1.5 ms). The dashed line indicates the threshold criterion (1%). Units with violation rates below 1% were classified as high-quality units (*n*=122).

stimulus intervals and fewer repetitions was used. Units were regarded as channel-level extracellular spiking activity, mainly multi-unit activity (MUA), and were retained only when they satisfied the following reliability and quality requirements.

512 candidate recorded units were initially obtained across all recording sessions and channels. Channels showing unstable recordings, such as excessive noise fluctuation or frequent signal saturation, were removed. Units whose stimulus-evoked firing rates failed to exceed baseline firing rates by at least 20% were also excluded to avoid unreliable estimates of selectivity. Second, to confirm response reliability, units were considered reliable when the inter-trial response correlation across the three repetitions was greater than 0.3 (Yoshida & Ohki, 2020). For each reliable unit, responses were averaged across repetitions to generate a mean response profile across 300 stimulus conditions. Application of these criteria resulted in 253 reliable units, which were included in the main analyses. Third, to examine whether the main findings were resistant to possible unit mixing, isolation quality was evaluated using refractory-period violations: for each unit, the fraction of spikes with ISI<1.5 ms was calculated. A conservative high-quality subset was defined using a violation rate<1% (122 units), and the principal selectivity and bias analyses were repeated for this subset. Finally, the resulting response vectors were normalized to a 0–1 range using min-max scaling.

Neuronal tuning properties

From the complete set of 200 images comprising 8 object categories (25 images per category), neural responses were arranged into a response matrix *R*. To estimate the preference of each neuron for specific categories, the category selectivity index (CSI) was calculated, which reflects how strongly a neuron responds to a subset of categories compared with the remaining categories. The CSI ranges from 0 to 1, with larger values representing stronger selectivity toward fewer categories. It was defined as:

$$\mathrm{CSI} = \frac{R_{max} - \bar{R}}{R_{max} + \bar{R}} \tag{1}$$

where R_{max} represents the maximum average response across the 8 categories, and $\bar{R}$ represents the mean response across all categories.

To evaluate how effectively the responses of a neuron separate different categories, the separation ratio (SR) was also calculated, which describes the ratio of inter-category variance to intra-category variance. A larger SR indicates stronger discriminability among categories. It was defined as:

$$\mathrm{SR} = \frac{\sigma^2_{between}}{\sigma^2_{within}} \tag{2}$$

where $\sigma^2_{between}$ indicates the variance of mean responses across categories, and σ^2_{within} indicates the average variance within each category.

To measure viewpoint selectivity, neural tuning curves derived from responses to 108 images (18 viewpoints×6 objects) were analyzed. For each neuron, several metrics were calculated to describe tuning depth and tuning regularity across viewpoints.

The viewpoint selectivity index (VSI) was used to estimate the degree to which responses were concentrated on a subset of viewpoints. For each object, VSI was calculated from the 18-dimensional viewpoint response vector using a selectivity form similar to CSI:

$$\mathrm{VSI} = \frac{R_{max} - \bar{R}}{R_{max} + \bar{R}} \tag{3}$$

where R_{max} represents the maximum average response across the 18 viewpoints, and $\bar{R}$ is the mean response across all viewpoints. The final VSI for each neuron was calculated by averaging the object-wise VSI values across the 6 objects.

Modulation depth (MD) was applied to quantify the dynamic range of neuronal responses across different viewpoints. This measure reflects the sensitivity of a neuron to changes in viewing angle and was defined as:

$$\mathrm{MD} = R_{max} - R_{min} \tag{4}$$

where R_{max} and $R_{\min}$ represent the maximum and minimum responses across all viewpoints, respectively.

Neighbor correlation (NC) measures the smoothness of the tuning curve by determining the correlation between adjacent viewpoint responses. It was defined as:

$$\mathrm{NC} = \frac{1}{2}\left[corr\left(r, r_{left}\right) + corr\left(r, r_{right}\right)\right] \tag{5}$$

where r is the 18-dimensional response vector, r_{left}, and r_{right} represent the circularly shifted vectors to the left and right, respectively. If the correlation was negative, it was set to zero.

To determine whether a neuron preferentially encoded category information or viewpoint information, the Tuning bias index (TBI) was defined as:

$$\mathrm{TBI} = \frac{TI_{class} - TI_{view}}{TI_{class} + TI_{view}} \tag{6}$$

where TI_{view} represents the mean class-tuning index (computed as $(R_{max} - R_{min}) / (R_{max} + R_{min})$), and TI_{class} denotes the mean view-tuning index for each object. TBI>0 indicates preferential category tuning, TBI<0 indicates preferential viewpoint tuning, and TBI=0 suggests comparable tuning for both dimensions.

Feature-fitting analysis

To estimate how strongly ENTO responses could be explained by particular visual attributes, multiple sets of image features were extracted for each stimulus, and the mean neuronal response across all stimulus conditions was fitted using feature-based regression models. The feature sets included color, shape, texture, V1-like and V2-like features, and deep convolutional features.

Color features: For each image, color statistics within the object region were calculated using the provided binary mask, with the background excluded. Color descriptors included (i) RGB and HSV histograms (50 bins per channel; normalized to sum to 1), (ii) cumulative histograms for each channel, and (iii) color moments (1st–3rd order: mean, standard deviation, and cubic root of the third absolute central moment) for each RGB/HSV channel. All color descriptors were merged into one color feature vector.

Shape features: Based on the object mask, the largest connected component was identified, and standard region-shape descriptors were extracted, including area, major/minor axis lengths (from the second central moments), aspect ratio (minor/major), eccentricity, orientation, and the seven Hu moments (scale/rotation/translation invariant). These descriptors were combined to form a shape feature vector.

Texture features: Texture was calculated from grayscale images within the object region. The extracted texture features included (i) gray-level co-occurrence matrix (GLCM) statistics over distances {1 2, 4, 8} and angles {0, 45, 90, 135} degrees

(contrast, dissimilarity, homogeneity, energy, correlation, and ASM), (ii) a uniform LBP histogram (radius =1, 8 sampling points; normalized), and (iii) steerable pyramid subband statistics (mean, standard deviation, and energy for each subband) obtained using steerable pyramid decomposition. All texture descriptors were merged into a single texture feature vector.

Model-based visual features: In addition to hand-crafted low-level descriptors (color/shape/texture), features were extracted from a hierarchy of computational vision models to represent increasingly complex visual information. Specifically, V1-like features (Pinto et al., 2008) were computed using a bank of oriented filters and local pooling operations, whereas V2-like features (Freeman & Simoncelli, 2011) were derived using a texture/summary-statistics representation. Deep convolutional features were also extracted from AlexNet (Krizhevsky et al., 2017), using activations from layers Conv1–Conv5 as progressively higher-level feature spaces. For each model/layer, activations were flattened into a feature vector, standardized across stimuli (z-scored), reduced using PCA (95% variance retained), and then entered into the same linear SVR fitting procedure described above. To evaluate whether the conclusions were robust across model choices and architectures, a broader set of feature extractors was further compared in Supplementary Figure S4, including AlexNet, VGG16, ResNet50, and InceptionV3, together with an untrained AlexNet and a transformer-based model, ViT-B/16.

Each feature set was standardized across stimuli by z-score transformation. To reduce dimensionality and limit overfitting, PCA was applied to each feature type, and components explaining 95% of the variance were retained. For each neuron, a linear support vector regression (SVR) model was trained to predict the mean firing rate of the neuron from the corresponding feature vector. Model performance was assessed using five-fold cross-validation, and goodness of fit was measured as the Pearson correlation coefficient between predicted and observed responses. A higher r value indicates that the feature set more effectively explains the response pattern of the neuron. For reproducibility, the complete feature-extraction and fitting pipeline has been released at https://github.com/BrainIOLab/ENTO-Representation.git.

Neuronal population analysis

To reduce potential biases caused by uneven distributions of viewpoint preferences, we constructed "viewpoint-balanced" neuronal subpopulations based on the tuning center of each unit. Specifically, the full 0–360° azimuthal space was equally divided into V=18 bins, from which k=2 neurons were randomly sampled per bin to generate a viewpoint-balanced subpopulation of n=36 neurons. For each object category, this sampling procedure was repeated N_{rep}=100 times, yielding 100 subpopulations for statistical and performance evaluation. If a given bin contained fewer than k neurons, neurons from neighboring bins were used as substitutes.

To evaluate the capacity of neuronal populations to discriminate object categories or identities, linear SVM classifiers were trained. The response vector of each neuron was z-score normalized and entered as input, with image category or identity used as the output label. Classification performance was assessed using five-fold cross-validation, and the mean accuracy was reported. To avoid artificial improvement of performance caused by repetition leakage, neural responses were first averaged across repeated presentations of the same stimulus condition, and cross-validation was then conducted on these condition-level samples (i.e., repetitions of the same stimulus were never separated between training and test folds). To examine how decoding performance changed with neuronal population size, subsets of neurons were randomly sampled from the full pool, and decoding accuracy was averaged across 20 repetitions for each size.

MDS was applied to the population response vectors (253 neurons per stimulus) or image feature vectors, projecting them into a two-dimensional space. This approach enabled intuitive visualization of category clusters and viewpoint trajectories. Representational similarity analysis (RSA) was conducted on the original data without dimensionality reduction (Kriegeskorte et al., 2008). For each stimulus set, representational dissimilarity matrices (RDMs) were constructed, in which each entry represented the dissimilarity between population responses to a pair of stimuli, calculated as 1−Pearson correlation. Comparable RDMs were generated from different image feature spaces using the same dissimilarity metric. To evaluate the correspondence between neural and feature-based representations, Spearman rank correlations were calculated between the upper triangular elements of the neural and feature RDMs. Statistical significance was determined using permutation tests.

To quantify how neuronal populations encode viewpoint-related geometric organization, two complementary metrics were defined: procrustes circularity deviation (PCD) and global discriminability (GD). PCD measures the deviation of 18 viewpoint representations from an ideal circular manifold embedded in representational space. A smaller PCD value indicates that viewpoint representations form a more continuous and geometrically regular ring-like structure. For a given object, the neural response vector of the 18 viewpoints can be expressed as:

$$Y = [y_1, \ldots, y_{18}]^{\mathrm{T}} \in \mathbb{R}^{18\times k} \tag{7}$$

where k denotes the number of neurons. The ideal unit circle template is defined as:

$$Y^{*} = [[\cos\theta_v, \sin\theta_v]]_{v=1}^{18}, \theta_v = \frac{2\pi}{18}(v-1) \tag{8}$$

Using Procrustes analysis, the empirical responses Y were aligned to the ideal template Y^{*} by optimizing a rotation matrix R, translation vector v, and scaling factor s, while minimizing the following cost function:

$$\min_{s,\boldsymbol{R},t} \frac{1}{V}\sum_{v=1}^{V} \boldsymbol{y}_v - \left(s\boldsymbol{R}\boldsymbol{y}_{\boldsymbol{v}}^{*} + t\right)_2^2, V = 18 \tag{9}$$

the minimum value of this alignment error was taken as the PCD, which measures how closely the neural manifold approaches a perfect circular structure.

The GD metric measures the overall separability of viewpoint representations by calculating the mean pairwise Euclidean distance between all viewpoint samples in the embedding space. A higher GD reflects stronger representational distinction among viewpoints. It is defined as:

$$\mathrm{GD} = \frac{2}{V(V-1)} \sum_{1\le i\le j\le V} y_i - y_{j2} \tag{10}$$

To visualize the geometric structure of high-dimensional population responses, classical multidimensional scaling (MDS) was applied (Borg & Groenen, 2005). MDS projects

stimuli into a lower-dimensional space (typically 2D) while preserving pairwise dissimilarities as closely as possible. In this analysis, RDMs were generated using correlation distance (1−Pearson correlation) between population response vectors. Classical metric MDS was then applied to the RDM to obtain a two-dimensional embedding that retained the global distance structure.

Neural tuning model simulation

To mechanistically examine how single-neuron tuning parameters shape population manifold geometry and representational performance, a parameterized joint tuning model over category and viewpoint was constructed, and the independent effects of category-tuning strength and viewpoint-tuning width were systematically scanned in a two-dimensional parameter space.

Let the number of object categories be *O*=6 and the number of viewpoints be *V*=18. Viewpoints were uniformly sampled on a circular domain as

$$\theta_v = \frac{2\pi}{V}(v-1), v \in \{1,\ldots,V\} \tag{11}$$

We simulated *N*=18 neurons responding to all (o,θ_v) conditions, forming a response matrix $\mathbf{R} \in \mathbb{R}^{N\times(OV)}$. Each neuron was assigned a preferred category $\mu_n^{(c)} \in \{1,\ldots,\mathrm{O}\}$ and a preferred viewpoint $\mu_n^{(v)} \in \{1,\ldots,V\}$. To avoid geometric bias caused by uneven distributions of tuning centers, preferred viewpoint centers were arranged to cover 0–2π about uniformly (i.e., equal representation for each viewpoint center at the population level), thereby implementing balanced sampling. For each condition (o,θ), the neuronal response was determined by a category-tuning term and a viewpoint-tuning term. Category tuning was defined over a categorical distance $d_c\left(o,\mu_n^{(c)}\right)$, and viewpoint tuning was defined over the circular angular distance

$$d_\theta(\theta,\phi) = \min\left(|\theta-\phi|, 2\pi-|\theta-\phi|\right) \tag{12}$$

Both tuning components were modeled using circular-Gaussian forms:

$$g_c(o|\mu_n^{(c)},\sigma_{cls}) = \exp\left(-\frac{d_c\left(o,\mu_n^{(c)}\right)^2}{2{\sigma_{cls}}^2}\right) \tag{13}$$

$$g_v(\theta|\mu_n^{(v)},\sigma_{view}) = \exp\left(-\frac{d_\theta\left(\theta,\theta_{\mu_n^{(v)}}\right)^2}{2{\sigma_{view}}^2}\right) \tag{14}$$

These components were then combined additively to generate the final response, followed by normalization across conditions to match the empirical analysis pipeline):

$$\mathrm{r}_n(o,\theta) = \frac{1}{2}\left[g_c(o|\mu_n^{(c)},\sigma_{cls}) + g_v(\theta|\mu_n^{(v)},\sigma_{view})\right] \tag{15}$$

Here, σ_{cls} controls category-tuning strength, mainly modulating the single-neuron category selectivity index (CSI), whereas σ_{cls} controls viewpoint-tuning width, systematically affecting tuning smoothness, NC, and viewpoint selectivity, VSI. At the population level, σ_{view} was treated as OV condition vectors in an *N*-dimensional representational space, and the same dimensionality-reduction and metric computations as in the empirical analyses were applied: Isomap was used to embed condition vectors into 3D to visualize category clusters and viewpoint trajectories, and inter-class distance, intra-class distance, and procrustes circularity deviation (PCD; quantifying the deviation of each within-category viewpoint trajectory from an about circular geometry, with lower PCD indicating higher continuity and a more regular ring-like manifold) were calculated. In addition, linear classification with *k*-fold cross-validation was performed to obtain category decoding accuracy Acc, and the two-dimensional grid $(\sigma_{cls},\sigma_{view})$ to obtain a performance landscape. Within the high-decoding regime (Acc > 0.9), representative solutions were selected by maximizing intra-class distance, maximizing inter-class distance, or minimizing PCD, and their corresponding 3D embeddings were visualized to illustrate the characteristic geometries produced by different tuning parameter combinations.

RESULTS

Category and viewpoint tuning of ENTO neurons

We recorded 253 reliably responsive neurons from the ENTO of anesthetized pigeons across 16 fixation sessions (Figure 1A). During recordings, pigeons were shown 200 category stimuli (8 categories×25 images) and 108 viewpoint stimuli (6 objects×18 viewing angles), and neuronal responses were obtained from each unit. Stimuli were selected from the Caltech-256 dataset (Griffin et al., 2007) and custom 3D-rendered images (Supplementary Figure S1). Before formal stimulus delivery, sparse-noise mapping was performed to estimate the spatial coverage of each neuron's receptive field. ENTO neurons generally showed large receptive fields that covered a substantial part of the display (~110° visual angle). After coarse localization of the receptive field, test images were additionally presented at several candidate positions within the receptive field, and the position producing the strongest response was chosen as the preferred stimulus location for that neuron (Supplementary Figure S5). To reduce adaptation effects reported in previous experiments (Wang et al., 2024), long inter-trial intervals were used, and each stimulus was restricted to three repetitions. Neuronal responses were measured as the mean firing rate during a 50–500 ms window after stimulus onset, with baseline correction relative to pre-stimulus activity. Neurons were defined as reliable when trial-to-trial correlations across the three repetitions were greater than 0.3 (Yoshida & Ohki, 2020). Final response values were averaged across trials.

We first evaluated category tuning in ENTO neurons. Figure 1C presents normalized responses across all recorded sites, arranged according to increasing category selectivity. The population showed a gradual continuum from broadly responsive neurons to neurons with sharp category preference. Neurons with weak selectivity (Figure 1D) responded reliably after stimulus onset but showed largely overlapping firing rates among categories. By contrast, highly selective neurons (Figure 1E) exhibited clear preference for a small number of categories, with well-separated response profiles. To measure category discriminability, the separation ratio was calculated, with higher values indicating stronger differentiation between categories. The selective neuron shown in Figure 1E had a separation ratio of 0.87, whereas the broadly tuned neuron in Figure 1D had a value of 0.16. Across the population, the mean separation ratio was 0.36±0.17 (mean±s.d.; Figure 1H), suggesting that most ENTO neurons had moderate category selectivity, although a subset displayed much stronger categorical separability.

Next, viewpoint tuning in ENTO neurons was examined. For each object, neurons were separately sorted by their preferred viewing angle (Figure 1F), showing that preference centers

together spanned almost the full 0–360° range. Notably, the preferred viewing angles differed across objects within the same neuron, indicating that viewpoint tuning was object-dependent. To characterize viewpoint tuning properties (Figure 1G), two complementary metrics were used. Modulation depth (MD) describes the dynamic range of neuronal responses across viewpoints and reflects the possible contribution of a neuron to fine-scale viewpoint discrimination. Across the population, the mean modulation depth was 0.28±.09 (Figure 1I), indicating moderate viewpoint selectivity in most neurons. Neighbor correlation (NC) measures the smoothness of the tuning curve across adjacent viewing angles. Although some neurons showed strong modulation, low NC values reflected irregular or discontinuous tuning patterns (Figure 1G, bottom). The mean NC was 0.62±0.17 (Figure 1J), indicating that ENTO neurons generally showed smooth and continuous viewpoint tuning.

A time-resolved analysis (20 ms window, 10 ms step) was conducted to determine whether these measures exhibited different temporal profiles (Figure 2A). The results revealed broadly similar time courses across measures, with peak values occurring around 150–170 ms. This shared temporal pattern suggests that category and viewpoint tuning may emerge together within the same neurons. To examine this possibility, the relationship between single-neuron tuning properties across the two dimensions was analyzed (Figure 2B, C). Category selectivity measured using the viewpoint stimulus set ($CSI^{ViewSet}$) was significantly correlated with category selectivity measured using the category stimulus set (CSI^{CatSet}), and viewpoint selectivity ($VSI^{ViewSet}$) was positively associated with $CSI^{ViewSet}$. Importantly, these relationships were present both across all recorded units and within the high-quality subset defined by low refractory violation rates, showing that selectivity consistency was maintained across tasks.

To examine possible tuning preferences, a tuning bias index (TBI) was calculated. The population median TBI was –0.041, indicating that most neurons did not show a systematic preference for either dimension (Figure 2D). However, the two tails of the TBI distribution revealed a minority of units with strong category bias (CS units) or viewpoint bias (VS units). Finally, viewpoint tuning continuity (NC) was not significantly associated with VSI, suggesting that these two properties are largely independent (Figure 2E). Importantly, all these patterns were retained in the subset of well-isolated units defined by low refractory violation rates, indicating that the observed effects reflect intrinsic properties of ENTO neurons rather than artifacts caused by multi-unit contamination or spatial averaging.

Visual features explaining ENTO activity

To determine which visual features contribute to neural responses in ENTO, multiple features were extracted from each stimulus image, including color, shape, texture, V1-like features (Pinto et al., 2008), V2-like features (Freeman & Simoncelli, 2011), and hierarchical features from each convolutional layer of AlexNet (Krizhevsky et al., 2017) pretrained on ImageNet (Deng et al., 2009). All feature sets were normalized and then reduced using PCA, retaining components that accounted for 95% of the variance. For each

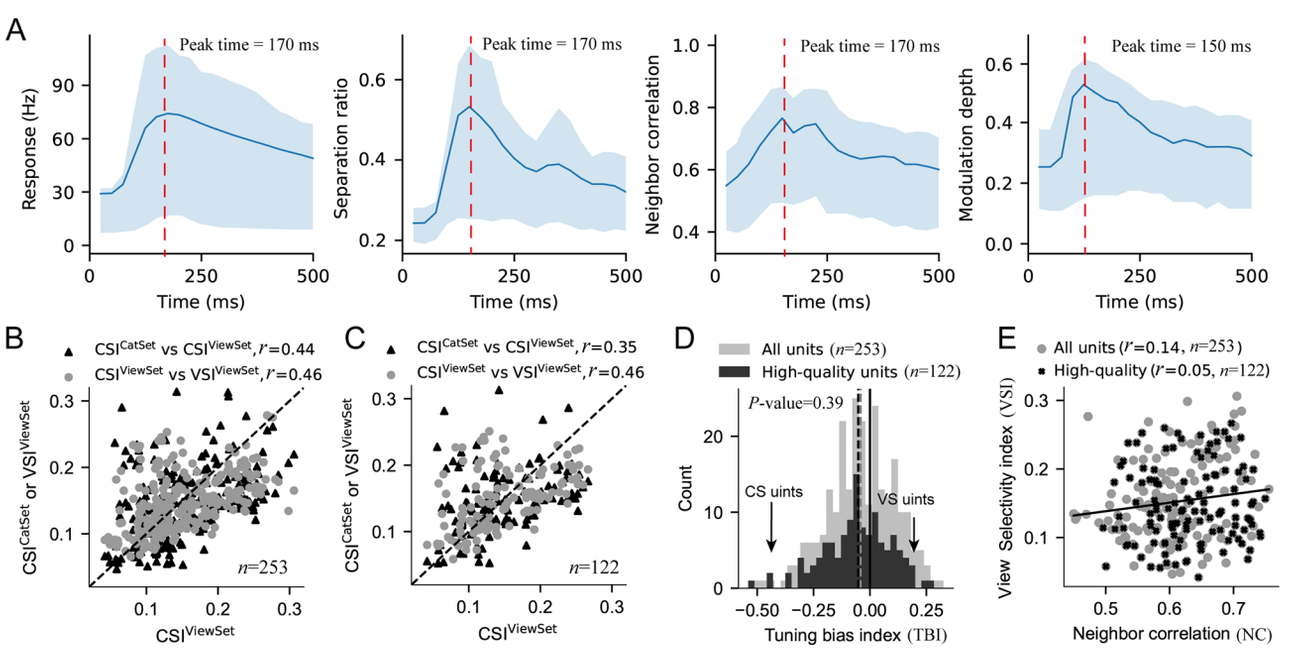

Figure 2 Population-level statistics of ENTO neurons

A: Time-resolved tuning parameters. Metrics were calculated using sliding windows (20 ms window, 10 ms step) across the first 500 ms following stimulus onset. Shaded regions represent the interquartile range (25th–75th percentiles). Red dashed lines indicate the peak latency of each metric. B: Relationships among selectivity indices across tasks for all recorded units (n=253 units). Black triangles indicate a significant positive association between category selectivity estimated from the category stimulus set (CSI^{CatSet}) and category selectivity estimated from the viewpoint stimulus set ($CSI^{ViewSet}$) (Pearson's r=0.44, R^2=0.19). Grey circles indicate a significant positive association between $CSI^{ViewSet}$ and viewpoint selectivity estimated from the viewpoint stimulus set ($VSI^{ViewSet}$) (Pearson's r=0.46, R^2=0.21). C: Relationships among selectivity indices across tasks for high-quality units (ISI violation<1%; n=122). D: Distribution of TBI values across units. Light grey bars correspond to all recorded units (n=253), whereas dark bars correspond to high-quality units defined by refractory violation rate<1% (n=122). No significant difference was detected between groups (Mann–Whitney U test, P=0.39). E: Correlation between NC and VSI across units. Grey circles represent all recorded units (n=253), and black crosses indicate high-quality units with low refractory violation rates (<1%; n=122). Pearson's r=0.14, R^2=0.02.

feature type, linear support vector regression (SVR) was applied to predict neuronal responses, with image features used as inputs and neuronal responses used as outputs (Figure 3A). The Pearson correlation coefficient between predicted and observed responses was used to quantify the goodness of fit for each neuron-feature pair.

Figure 3B summarizes the distribution of model fits across neurons for each type of visual feature. Among the low-level features, color produced the highest prediction accuracy, indicating that it is an important cue encoded in ENTO. In contrast, shape features produced the weakest fit, suggesting that shape alone cannot sufficiently explain neuronal responses. All convolutional layers of AlexNet showed relatively strong fits, with Conv5 reaching the highest median value (r=0.35). This pattern suggests that ENTO neurons may represent more abstract image information. In comparison, V1-like and V2-like features had lower explanatory power than color and deep convolutional features and were similar to texture features. These results imply that ENTO neurons are relatively less sensitive to local edge- and texture-like features that are typical of early visual areas.

To further examine whether this explanatory power reflected learned higher-level representations rather than architecture-related artifacts, an additional layer-wise analysis was performed across several model architectures (Supplementary Figure S4). For each neuron, its category-averaged response was fitted using SVR based on features extracted from successive layers. Trained CNNs consistently showed increased fitting performance with greater depth, whereas an untrained AlexNet did not show this depth-dependent improvement. These findings indicate that improved fits in deeper CNN layers are mainly driven by learned semantic-level representations that align more closely with ENTO category-related response structure.

Although most visual features produced only modest overall predictive performance, some neurons showed notably high fitting accuracy. For example, the best-fitting neuron for texture features reached a correlation of r=0.64 (Figure 3D). For these "best-matching" neurons, model predictions (red) closely followed the actual responses (black) across stimulus sequences (Figure 3D). The Top- 8 and Bottom- 8 stimuli that elicited the strongest and weakest responses in these neurons were visualized. For the color-optimal neuron (Figure 3E), the Top- 8 stimuli were mainly high-luminance images, whereas the Bottom- 8 stimuli were dominated by low-luminance images, suggesting sensitivity to brightness or luminance-

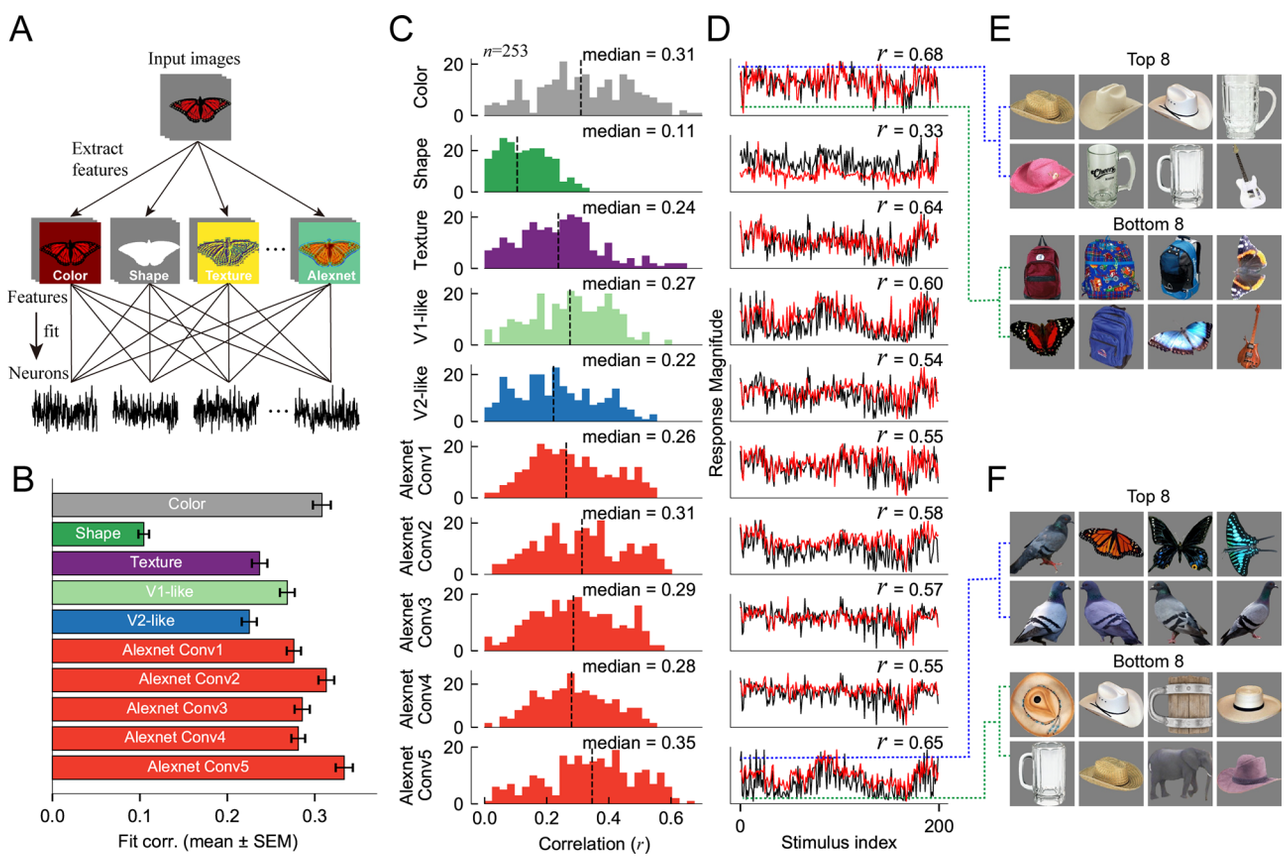

Figure 3 Explaining ENTO neuronal responses with image features

A: Overview of the analysis pipeline. For each stimulus, color, shape, texture, V1-like, V2-like, and AlexNet convolutional-layer features were extracted, normalized, and reduced using PCA while retaining 95% of the cumulative explained variance. Each feature set was then used to predict neuronal responses with a linear SVR model, and performance was evaluated by five-fold cross-validation. Pearson's correlation coefficient (r) between predicted and observed responses was used as the measure of goodness of fit. B: Population summary. Model-fitting performance across 253 neurons (mean±SEM). C: Distributions of model-fitting performance for each feature type. Dashed lines indicate median values. Although overall fitting performance was modest, a subset of neurons showed high predictability (right tail). D: Representative "best-matching" neurons for each feature type. Predicted responses (red) closely followed observed firing rates (black) across the stimulus sequence. Pearson's r indicates the goodness of fit. E: Color-optimal neuron. Top-8 and bottom-8 stimuli producing the highest and lowest responses, respectively. The highest-ranked stimuli were mainly composed of high-luminance images. F: AlexNet Conv5-optimal neuron. Top-ranked stimuli clustered around categories such as pigeons and butterflies.

related visual energy. For the AlexNet Conv5-optimal neuron (Figure 3F), the Top- 8 stimuli clustered around categories such as pigeons and butterflies, while the Bottom- 8 stimuli showed no obvious regular pattern, indicating that this neuron may encode abstract object- or category-level information. Overall, feature-based modeling showed that color and deep convolutional representations, particularly those from higher AlexNet layers, provided the best explanation of ENTO neuronal responses.

Population geometry and categorical representation of ENTO neurons

To investigate how visual information is arranged across the ENTO population, multidimensional scaling (MDS) was applied to the responses of all recorded neurons. The resulting representational space exhibited a distinct geometric organization (Figure 4A), with stimuli continuously distributed along two about orthogonal axes corresponding to hue and luminance. Several semantic object categories (e.g., pigeon, butterfly, cowboyhat) also occupied relatively distinct cluster-like regions. To estimate the contribution of color-related properties, image hue and brightness statistics were extracted and correlated with the MDS dimensions. Hue was associated with the first and second MDS dimensions (r=0.31 and r=0.45), whereas brightness was associated with the second dimension (r=0.42), indicating that color-related information strongly shaped the representational geometry of ENTO.

Robustness was further evaluated using nonlinear embedding methods (UMAP and t-SNE; Figure 4B). Across these methods, a simple regression model based on hue and brightness explained a considerable proportion of the variance in the 2D embedding coordinates (R^2=0.18–0.29). Importantly, residual decoding demonstrated that category decoding remained above chance after hue or brightness was regressed out from neural responses (Figure 4C), indicating that category-discriminative signals persisted beyond low-level color and brightness statistics. Together, these findings suggest that the representational structure in ENTO is jointly determined by low-level physical attributes and higher-level categorical features.

We next measured the discriminability of stimulus categories in the ENTO population using a linear SVM classifier (Figure 4D). Across the eight object categories, the mean decoding accuracy was 67%±2% (mean±s.d.), which exceeded models based on low-level visual features (e.g., color, texture, shape) but remained lower than the performance obtained using deep convolutional features. Category decoding accuracy differed substantially among classes (Figure 4E). Biologically or experientially salient categories, such as pigeons and faces, showed markedly higher decoding accuracy than other categories (see Supplementary Figure S6 for the average neural responses and within-category response variability). Importantly, this advantage could not be explained only by low-level image statistics. For example, in the pigeon category, decoding based solely on color features reached only ~35% accuracy, whereas decoding based on ENTO population responses reached ~83%, showing that ENTO activity contains category information beyond basic color, shape, or texture features. To further control color-related confounds, decoding was examined as a function of neuron number before and after hue or brightness was regressed out from the neural responses (Figure 4F). Accuracy increased with population size and remained above chance after residualizing by hue or brightness, with only a modest reduction compared with raw responses. Together, these results indicate that category information is distributed across the ENTO population and is not solely explained by low-level color or brightness statistics.

Representational similarity analysis (RSA) was then performed. Neural and feature RDMs were calculated using correlation distance and compared using Spearman correlation. At the population level, the RDM showed significant similarity to several feature spaces, including category structure, low-level visual features (color, shape, texture), and hierarchical features from AlexNet (Figure 4G; Supplementary Figure S7). Similarity to AlexNet features was higher in early layers, decreased in intermediate layers, and then increased again in the deep Conv5 layer, indicating that ENTO representations include both low-level visual statistics and higher-level categorical organization.

To further examine the relationship between low-level visual features and category-related organization, neurons were grouped according to their fitting strength to different visual features, and RSA and category decoding were re-evaluated within each subpopulation (Figure 4H). As fitting strength increased, RSA similarity to the corresponding feature increased for all feature types, confirming that these subpopulations progressively captured feature-consistent structure. Importantly, stronger feature bias did not reduce category decoding. Subpopulations aligned with color and shape features showed stable or slightly improved decoding performance, whereas texture-aligned subpopulations showed decreased performance. Together, these results indicate that low-level visual features and category-related information coexist within ENTO.

Viewpoint representation and tuning mechanisms in ENTO neurons

We examined whether the pigeon ENTO represents structured viewpoint information and which neuronal properties contribute to this organization. To quantify these properties, two metrics were defined. The first, Procrustes circularity deviation (PCD), measures how closely the arrangement of neural responses across 18 evenly spaced viewpoints resembles a circular geometry in reduced-dimensional space. Lower PCD values indicate stronger continuity and a more ordered circular viewpoint manifold (Figure 5A, face). The second, global discriminability (GD), measures the average pairwise representational distance, in reduced-dimensional space, between neural responses across viewpoints. Higher GD values reflect greater separability among views (Figure 5A, pigeon).

Considering possible biases caused by uneven distributions of tuning centers, balanced neuronal subpopulations were constructed. Specifically, for each of the 18 viewpoints, two neurons with a preferred viewpoint matching that angle were randomly selected. This produced a balanced subpopulation of 36 neurons (2 neurons×18 viewpoints), ensuring equal representation of tuning centers across viewpoints. The procedure was repeated 100 times (Supplementary Figure S8). For each subset, the mean modulation depth (MD) and neighbor correlation (NC) were calculated, and their associations with PCD and GD were examined. Overall, NC was significantly negatively correlated with PCD (Spearman's ρ=−0.39), whereas MD showed a weaker negative correlation (Spearman's ρ=−0.29; Figure 5B). By contrast, MD showed a strong positive correlation with GD (Spearman's ρ=0.95), while NC showed a moderate correlation (Spearman's ρ=0.77;

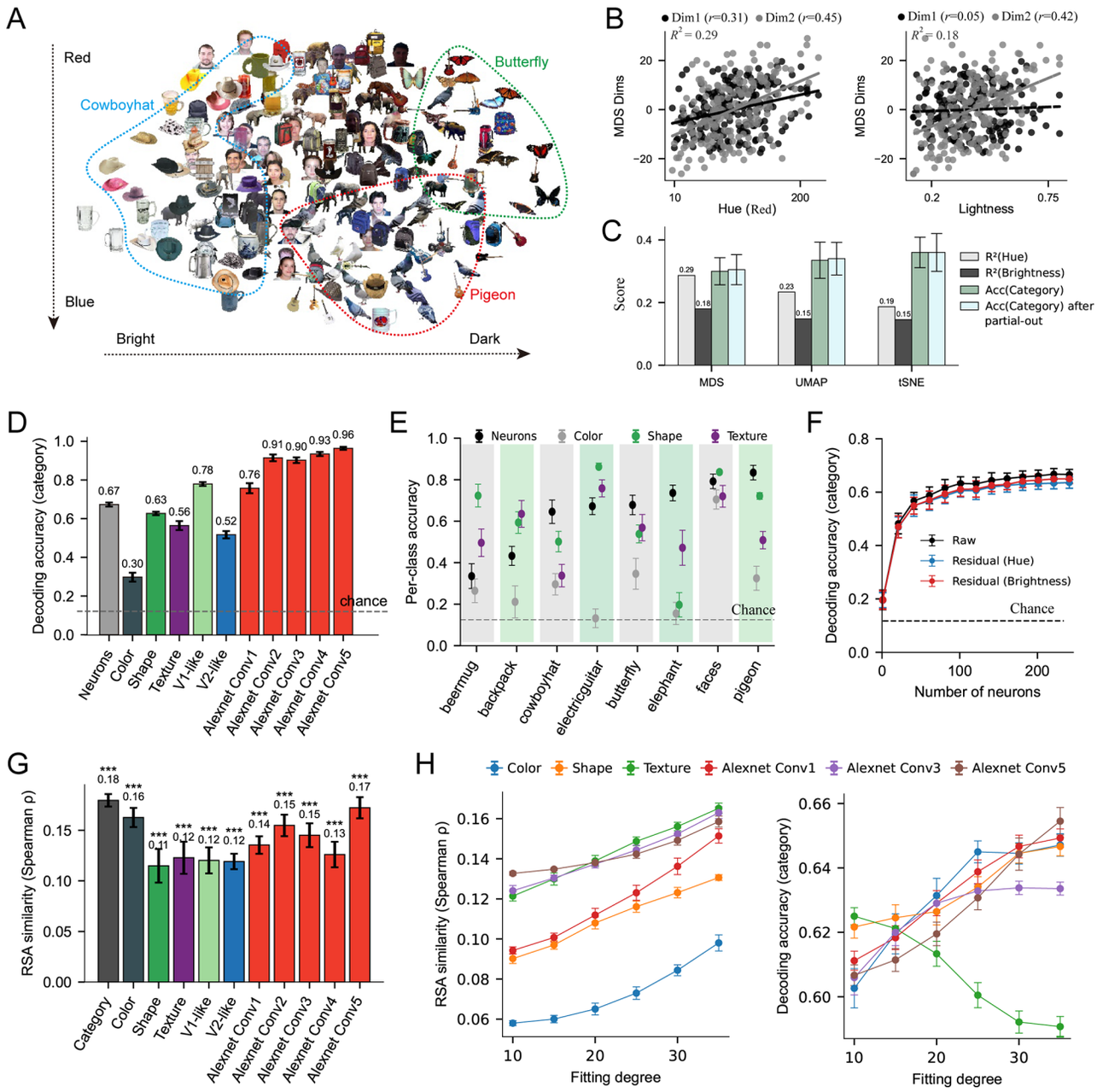

Figure 4 Population-level representational geometry and decoding performance of ENTO neurons

A: Two-dimensional MDS embedding of population responses across all recorded ENTO neurons (253 units×200 images). Colored dashed contours highlight representative category groupings within the embedded space. B: Relationship between MDS dimensions and stimulus color statistics. Scatterplots show MDS dim1/dim2 plotted against hue (foreground red-channel mean intensity) and lightness (foreground mean luminance), together with linear regression fits. Reported values are Pearson's *r* and overall R^2. C: Summary of feature representation in 2D embeddings. For each embedding method, hue R^2, brightness R^2, decoding accuracy, and decoding accuracy after removing hue and brightness contributions were quantified. Category decoding was performed using a linear SVM (5-fold CV). Bars represent mean±SD. D: Decoding performance across different feature spaces using a linear SVM (5-fold CV, 50 repetitions). Bars indicate mean±SD. E: Per-class decoding accuracy (recall) for ENTO responses and basic image features (color, shape, and texture). F: Decoding accuracy as a function of neuron number and residual controls. For randomly selected ENTO subsets of increasing size, linear SVM accuracy (5-fold CV, 50 random splits) is shown for raw responses and for responses after residualizing hue or brightness. Curves indicate mean±SD. G: Representational similarity analysis. Neural and feature RDMs (correlation distance) were compared using Spearman correlation of the upper triangular elements. Statistical significance was assessed with permutation testing. Bars show the observed ρ values. H: Subpopulation analysis based on fitting degree. After removing highly correlated neurons (pairwise correlation threshold was 0.90), neurons were ranked according to their single-neuron fitting score for each feature. Sliding subgroups of 80 neurons were evaluated across the sorted list. Left: RSA. Right: Decoding accuracy derived from subgroup responses. Error bars represent SEM across windows.

Figure 5C). These results indicate that both tuning parameters are closely related to viewpoint representation performance: NC mainly contributes to geometric circularity (lower PCD), whereas MD increases overall separability (higher GD). Moreover, stimuli such as faces and pigeons consistently produced lower PCD values under balanced sampling (Figure 5B), suggesting that ENTO neurons form more continuous and structured viewpoint representations for

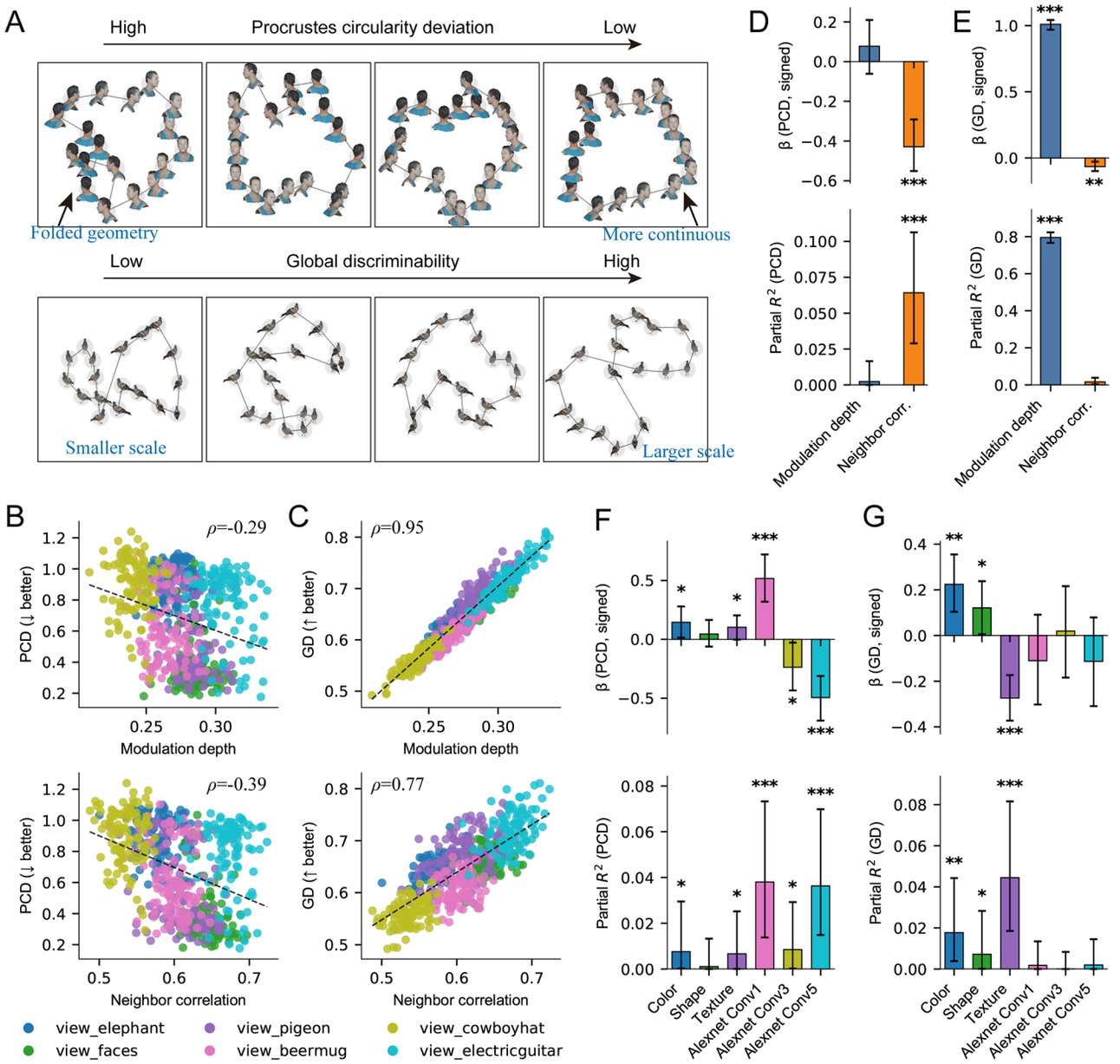

Figure 5 Viewpoint representation of ENTO neuronal populations

A: Schematic illustration of viewpoint-representation metrics. Top: Procrustes circularity deviation (PCD); smaller values indicate more circular and continuous neural manifolds. Bottom: Global discriminability (GD); larger values indicate stronger separation among viewpoints. B: Relationship between PCD and tuning properties (MD, NC). Neuronal subpopulations (n=36) were generated by sampling 2 neurons from each of 18 viewpoint bins to ensure balanced tuning centers, and the procedure was repeated 100 times. MD and NC were calculated for each subgroup. Different colors represent the six object categories. Correlations were evaluated using Spearman's ρ. C: Relationship between GD and tuning properties. D: Directional and independent contributions of tuning properties to PCD. Standardized multiple linear regression was used to estimate signed β-coefficients (top) and partial R^2 values (bottom). Error bars indicate bootstrap-derived 95% confidence intervals. Asterisks indicate significance levels: *: P<0.05, **: P<0.01, ***: P<0.001. E: Independent contributions of tuning properties to GD. F: Directional and independent contributions of feature tuning (fitting scores) to PCD. Multiple regression analyses were performed using neuronal fitting scores for color, shape, texture, and AlexNet Conv1/Conv3/Conv5 as predictors. G: Contributions of feature tuning to GD. The same analysis as in (F), except that feature tuning was regressed against GD.

behaviorally or ecologically salient categories.

To further separate the directionality and independent contributions of MD and NC to viewpoint representation, standardized multiple linear regression was conducted, and both the signed β coefficients and partial R^2 values were reported. For PCD, NC showed a stronger negative β and a substantially higher partial R^2 than MD (Figure 5D), indicating that, after controlling for MD, NC made a distinct and independent contribution to reducing PCD, that is, improving circularity. In contrast, MD showed a smaller β and lower partial R^2, suggesting only a limited direct effect on circular structure. For GD, MD showed the largest positive β and the highest partial R^2, whereas the independent effect of NC was weak (Figure 5E). These findings agree with the correlation analyses, showing that PCD is primarily driven by NC, which enhances continuity, while GD increases mainly with stronger MD, which enhances separability.

Next, we included the fitting degree of each neuron to

multiple visual features (color, shape, texture, and AlexNet Conv1/3/5) as predictors in separate regression models for PCD and GD. The results showed that neurons with stronger fits to AlexNet features (Conv1 and Conv5) generally exhibited larger absolute β values and higher partial R^2 values (Figure 5F). Specifically, Conv1 showed a positive β, whereas Conv5 showed a negative β, suggesting that higher-level abstract representations made a stronger and independent contribution to reducing PCD, that is, improving circularity. For GD, basic visual features (color, shape, texture) produced larger β and higher partial R^2 values (Figure 5G), indicating that low-level feature tuning more directly increased viewpoint separability. Together, these findings reveal complementary contributions of different feature levels: high-level abstract features mainly promote representational continuity (lower PCD), whereas low-level image attributes more directly enhance global discriminability (higher GD).

Decoupling category separability and viewpoint geometry

Viewpoint invariance describes the capacity to recognize an object despite continuous changes in viewing angle. To separate two major components of this process—category separability (the ability to distinguish object identity) and viewpoint geometry (the geometric arrangement of different views)—we quantified three complementary measures for both the complete ENTO population and tuning-balanced subgroups: (1) classification accuracy obtained using a linear support vector machine (SVM); (2) within-class and between-class distances as measures of category discriminability; and (3) Procrustes circularity deviation (PCD), which evaluates how closely viewpoint representations conform to an ideal circle.

Functionally, object identity could be decoded with high accuracy using only a relatively small subset of ENTO neurons. Nearly ceiling-level classification performance (~100%; Figure 6A) was achieved with as few as 20 neurons, and almost all balanced subgroups containing n=36 neurons reached accuracies greater than 95% (Figure 6B, C). These findings indicate that object recognition across viewpoint changes is highly efficient within the ENTO and does not depend on large-scale population recruitment. However, despite this strong decoding performance, MDS of the full neuronal population did not reveal a clear circular trajectory structure across viewpoints (Figure 6A). This observation suggests that viewpoint representation in the ENTO may depend on specific neuronal subsets rather than being uniformly distributed across the entire population.

To further investigate this issue, we generated 600 random neuronal subgroups, each consisting of n=36 neurons (Figure 6B, C; each dot represents one subgroup). Some of these subgroups exhibited low PCD values, indicating that certain neuronal subsets were capable of generating more continuous viewpoint trajectories. Despite these local effects, the relationships between within-class distance, between-class distance, and PCD remained weak. No significant association was found between PCD and category classification accuracy (Figure 6D). Likewise, correlations among all representational metrics were uniformly weak, with effect sizes close to zero across the 600 randomly sampled subgroups (Figure 6F–H). Collectively, these results demonstrate that category separability (high between-class distance and low within-class distance) and viewpoint continuity (low PCD) represent two largely independent dimensions of ENTO population coding.

Among the randomly generated subgroups, we identified subsets that simultaneously exhibited high category decoding accuracy and strong viewpoint continuity, as well as others that preferentially expressed only one of these properties. To illustrate these differences, MDS trajectories from three representative extreme subgroups were visualized (Figure 6E). Subgroups with larger between-class distances displayed more clearly separated category clusters, thereby providing greater representational capacity for encoding within-class structure. In contrast, subgroups with smaller within-class distances formed more compact category clusters, facilitating more precise classification. Finally, subgroups with lower mean PCD values generated viewpoint trajectories that more closely resembled an ideal circular arrangement, reflecting smoother and more orderly viewpoint encoding (as illustrated for pigeon, face, and beer mug). Notably, no individual subgroup consistently produced optimal viewpoint representations across all object categories, suggesting that different object classes may recruit distinct ENTO neuronal subpopulations.

Tuning mechanisms underlying category separability and viewpoint geometry

The preceding analyses demonstrate that ENTO population activity simultaneously supports strong category separability and varying degrees of viewpoint-related geometric continuity. Importantly, these representational properties are not evenly distributed across all neurons but instead differ among neuronal subsets. To account for the coexistence of category separation and viewpoint continuity in ENTO, we propose a tuning-driven subpopulation mechanism in which manifold geometry and representational performance are primarily determined by neurons with specific tuning characteristics rather than by uniform contributions from the entire population.

We first constructed functional subpopulations from empirically recorded ENTO neurons. Figure 7A illustrates the viewpoint-tuning curves of a representative neuron across six object categories and 18 viewpoints. To determine whether continuous geometry could arise trivially from assumptions of smoothness, the same tuning profiles were reconstructed using two contrasting models: a continuous Gaussian function and a discontinuous step function (Figure 7A). Three neuronal subpopulations were then defined by ranking neurons according to single-neuron metrics: NC-max (highest tuning continuity; neighbor correlation, NC), CSI-max (highest category selectivity index, CSI), and VSI-max (highest viewpoint selectivity index, VSI). In addition, two reconstruction-based control groups were generated using the same NC-max neurons: Gauss-sim(NC) and Step-sim(NC), derived from Gaussian and step-function reconstructions, respectively. As expected, the resulting metric distributions were clearly dissociated (Figure 7B). NC-max, CSI-max, and VSI-max preferentially maximized NC, CSI, and VSI, respectively. Gaussian reconstruction further increased NC, whereas step-function reconstruction reduced NC and simultaneously weakened VSI.

We next quantified representational performance and visualized the corresponding 3D embeddings for these five subpopulations (Figure 7C, D). The resulting structures closely reflected the tuning properties used to define each subgroup. NC-max generated the smoothest viewpoint trajectories and produced the lowest PCD values. CSI-max yielded stronger separation among category clusters, reflected by larger inter-class distances. VSI-max produced more extended within-

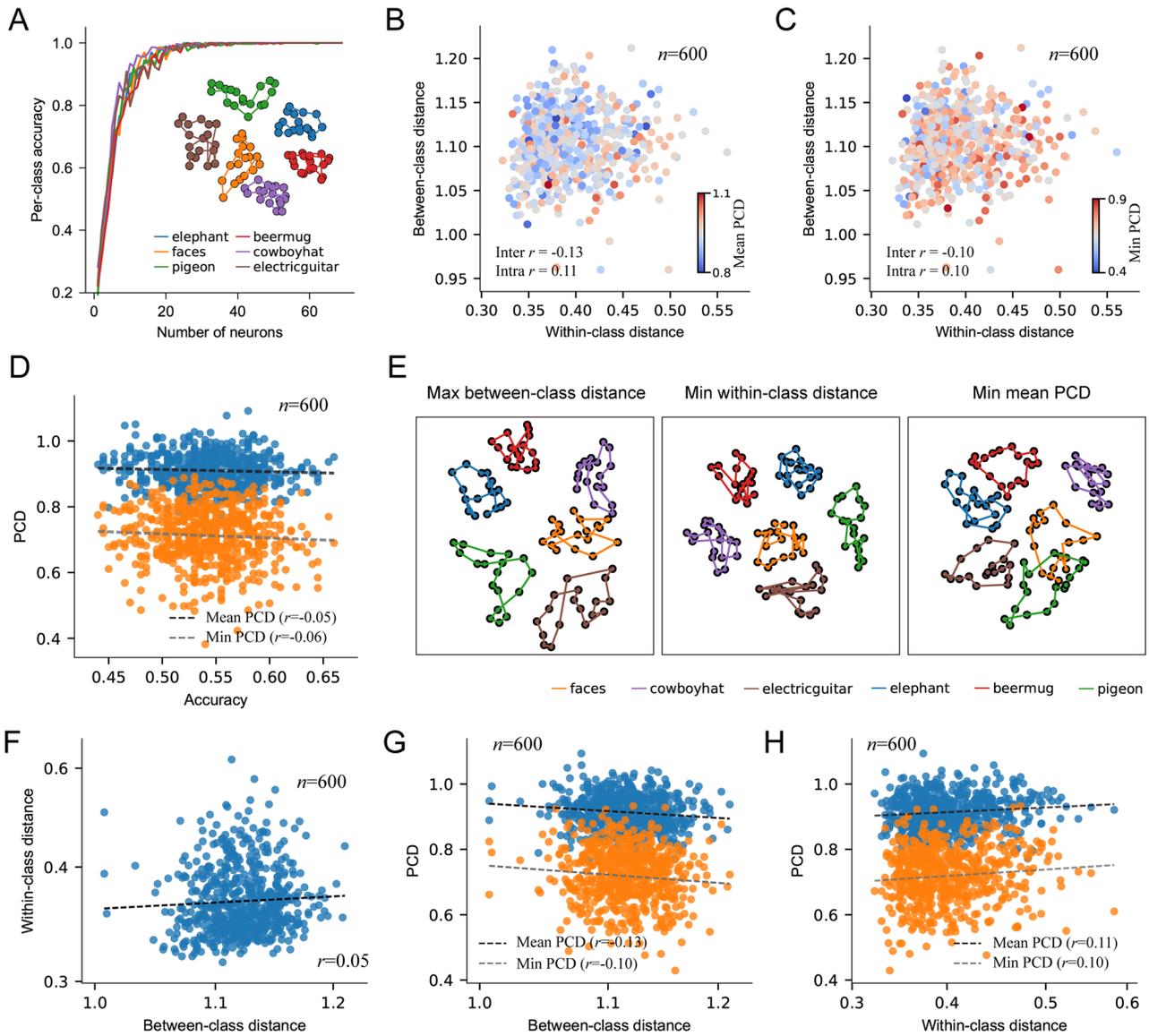

Figure 6 Viewpoint invariance and representational structure of ENTO neuron populations

A: Decoding accuracy as a function of neuron number. Different numbers of neurons were randomly sampled for linear SVM classification (5-fold CV, repeated 20 times). The curve represents mean accuracy. Inset: MDS embedding of the full neuronal population. B: Relationship between within-class/between-class distances and mean PCD. Under balanced sampling (two neurons per viewpoint; n=36), 600 random subgroups were generated. Each point represents a subgroup and is colored according to mean PCD (blue=low, red=high). Pearson's r was calculated. C: Relationship between within-class/between-class distances and minimum PCD. The same analysis as in (B), using the minimum PCD value within each subgroup. D: Correlation between viewpoint circularity and category decoding accuracy. Blue and orange points correspond to mean and minimum PCD, respectively; Pearson's r was calculated. E: Representative MDS trajectories for three subgroups. Left: subgroup with the largest between-class distance. Middle: subgroup with the smallest within-class distance. Right: subgroup with the lowest mean PCD. F: Correlation between within-class and between-class distances. Each point represents one of 600 randomly sampled neuronal subgroups (n=36; 2 neurons per viewpoint bin). Linear regression yielded Pearson's r=0.05, R^2=0.0025. G: Correlation between PCD and between-class distance. Pearson's r=0.13, R^2=0.017 for mean PCD, and r=0.10, R^2=0.01 for min PCD. H: Correlation between PCD and within-class distance. Pearson's r=0.11, R^2=0.012 for mean PCD, and r=0.10, R^2=0.01 for min PCD.

category trajectories, resulting in larger intra-class distances. Importantly, simply altering the fitting function substantially changed the geometry of the representation. Gaussian reconstruction further enhanced tuning continuity and reduced PCD, whereas step-function reconstruction markedly disrupted continuity and reduced intra-class distance. Together, these findings demonstrate that population-level manifold geometry is systematically shaped by the distribution of tuning properties at the single-neuron level.

To mechanistically evaluate how tuning parameters influence manifold geometry, we developed a parameterized tuning model (Figure 7E). In this model, 18 neurons were simulated to respond to six objects across 18 viewpoints using circular Gaussian tuning functions, and two independent parameters were varied: σ_{cls}, which controls category-tuning strength and primarily modulates CSI, and σ_{view}, which controls viewpoint-tuning width and mainly affects NC and VSI (Figure 7F). We then calculated intra-class distance, inter-class distance, and PCD across the two-dimensional parameter space (σ_{cls}, σ_{view}) (Figure 7G). These maps revealed a clear performance landscape, in which distinct regions preferentially maximized inter-class separation,

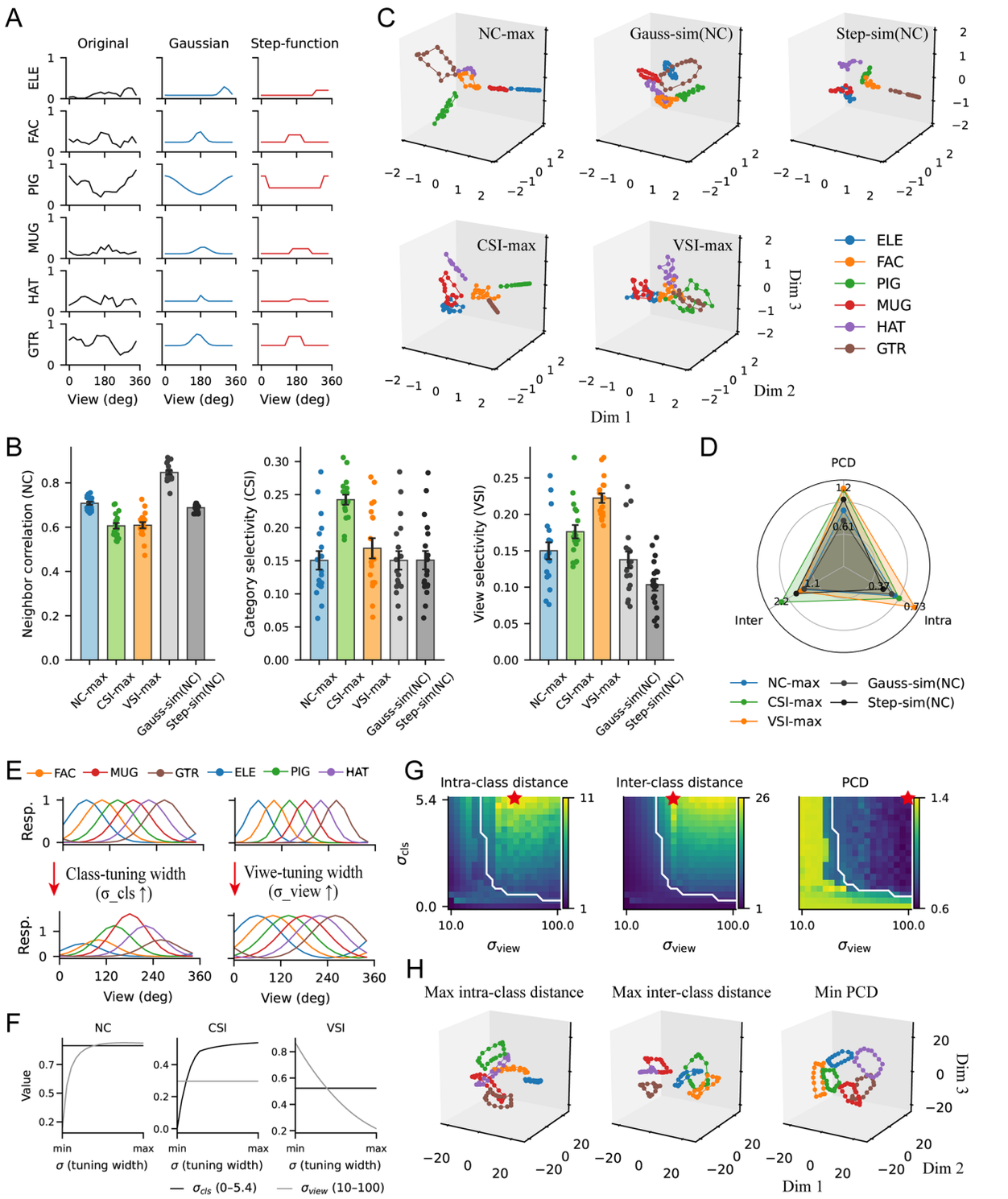

Figure 7 Tuning-dependent formation of population manifolds in ENTO

A: Example viewpoint-tuning curves of a representative neuron across 6 object categories (ELE, elephant; FAC, faces; PIG, pigeon; MUG, beer mug; HAT, cowboy hat; GTR, electric guitar) and 18 viewpoints (0–360°). Original responses are compared with reconstructions generated using smooth Gaussian functions and discontinuous step functions. B: Distributions of single-neuron tuning metrics across five subpopulations: NC-max (highest neighbor correlation, NC), CSI-max (highest category selectivity index, CSI), VSI-max (highest viewpoint selectivity index, VSI), and two reconstruction-based controls derived from the same NC-max neurons: Gauss-sim(NC) and Step-sim(NC). Bars represent mean±SEM. C: Three-dimensional Isomap embeddings of population responses for each subpopulation. D: Summary comparison of representational performance across subpopulations, quantified by inter-class distance (Inter), intra-class distance (Intra), and PCD. E: Parameterized tuning simulation showing independent manipulation of category-tuning strength (σ_{cls}) and viewpoint-tuning width (σ_{view}). F: Effects of varying σ_{cls} or σ_{view} on population-averaged NC, CSI, and VSI in the simulation. G: Performance landscape across the two-dimensional parameter space (σ_{cls}, σ_{view}). Heatmaps show intra-class distance, inter-class distance, and PCD. White contours indicate decoding accuracy greater than 90%. Red stars mark representative optima corresponding to maximal intra-class distance, maximal inter-class distance, and minimal PCD. H: Three-dimensional embeddings corresponding to the representative optima highlighted in (G).

expanded within-class viewpoint structure, or minimized PCD to generate near-circular manifolds. Restricting the analysis to the high-decoding regime (white contour; accuracy >90%), three representative solutions located at the extrema of these objectives were selected, and their characteristic geometries were confirmed in the corresponding embeddings (Figure 7H). Overall, these results support a subpopulation-based coding strategy in the ENTO, in which neurons function as locally tuned dictionary elements across category and viewpoint dimensions. Sparse or structured combinations of these elements can generate low-dimensional manifolds that simultaneously support category separability and continuous viewpoint geometry.

DISCUSSION

Previous studies have suggested that the pigeon ENTO does not encode category information and instead responds mainly to low-level visual features (Anderson et al., 2020; Azizi et al., 2019; Clark et al., 2022). However, by using a broader stimulus set, we found that ENTO neurons do show category selectivity. In an eight-category decoding task, the ENTO population reached a decoding accuracy of ~67% (Figure 4D), and this performance could not be explained by low-level image properties alone, because models based on color, shape, or texture consistently produced much lower accuracy (Figure 4E, F). MDS of population responses revealed an organized representational geometry, with distinct category clusters appearing in neural space (Figure 4A). Notably, under our behaviorally relevant stimulus paradigm, ENTO ensembles showed particularly high decoding accuracy for faces and pigeons (Figure 4E). Importantly, the stronger discriminability of ENTO for behaviorally salient categories is consistent with the idea that neural systems may be predisposed to prioritize ecologically important signals. Converging evidence from developmental neuroscience also supports this possibility, showing that category-selective representations can appear early, even before extensive visual experience. For example, recent electrophysiological studies in visually naïve domestic chicks have identified neurons selectively tuned to faces (Kobylkov et al., 2024), suggesting that the encoding of behaviorally relevant visual categories may be shaped, at least partly, by innate predispositions in addition to postnatal learning (Costalunga et al., 2024; Lorenzi et al., 2025).

Consistent with previous reports (Hsiao et al., 2020; Niu et al., 2024), a strong influence of color on ENTO neuronal responses was observed. This was reflected by the higher feature-fitting scores for color features (Figure 3B) and the dominant color gradient present in the population-level MDS space (Figure 4A–C). Nevertheless, subgroup analyses indicate that the encoding of basic visual attributes is not mutually exclusive with categorical representation in ENTO: neurons with stronger fits to color and shape still supported robust category decoding (Figure 4H). This pattern resembles that of the IT cortex, where RSA analyses have revealed category-structured representations in visual cortex (Kriegeskorte et al., 2008), similar to the pattern observed in Figure 4G. At the same time, low-level visual information is not discarded but retained and progressively integrated along the ventral visual stream (Hong et al., 2016; Li et al., 2009). Moreover, both V4 and IT neurons have shown hue-invariant responses, supporting stable object recognition under changing chromatic conditions (Bushnell & Pasupathy, 2012; Chang et al., 2017). Taken together, these findings suggest that ENTO may likewise integrate low-level visual features (e.g., color and shape) to support partially invariant, category-discriminative representations.

Our results indicate that the ENTO supports efficient, viewpoint-invariant object recognition. Using only 20 neurons, the system could classify six object categories with high accuracy (Figure 6A). Meanwhile, a relatively high population-level NC was observed in ENTO (Figure 6B), suggesting that viewpoint tuning shows a certain degree of smooth continuity at the population level. Conceptually, this finding is consistent with the behavioral studies of Wood and Wood et al. in newly hatched chicks (Wood, 2013; Wood & Wood, 2015), which showed that animals exposed to objects undergoing continuous spatiotemporal transformations can form stable object-identity recognition and generalize to novel viewpoints. Previous studies have also shown that visual experience can reshape neuronal selectivity through both supervised and unsupervised mechanisms (Freedman et al., 2006; Li & DiCarlo, 2010; Matteucci & Zoccolan, 2020). It is hypothesized that, through visual experience, ENTO may acquire a set of foundational codes for object category and viewpoint. By flexibly combining these neurons, the system may construct neural manifolds that encode not only viewpoint but also category and other invariant object dimensions (Hashimoto & Kurata, 2000; Stringer et al., 2006).

Our findings show that manifold structure in ENTO does not emerge simply from modeling assumptions but is shaped by the distribution of tuning properties across neuronal subpopulations. Analyses of real data showed that subpopulations maximizing continuity, category selectivity, or viewpoint selectivity displayed systematic differences in embedding space (Figure 7C). Moreover, disruption of tuning smoothness markedly weakened manifold continuity (Figure 7D). Parameterized analyses further demonstrated that category-tuning strength and viewpoint-tuning width jointly determine population geometry and, within specific parameter regimes, can simultaneously support high category separability and smooth viewpoint continuity (Figure 7G). Overall, these results support a subpopulation-based, basis-function-like coding scheme (Bernardi et al., 2020; Rigotti et al., 2013), in which neurons function as locally tuned elements whose sparse and structured combinations generate low-dimensional neural manifolds that reconcile separation and continuity within a unified representational space.

To interpret the response properties of ENTO neurons, neuronal activity was modeled using image-derived visual features and hierarchical features extracted from AlexNet layers. Neurons with higher fitting scores consistently captured these key feature dimensions (Figure 3D–F). Among the trained CNNs, fitting performance generally increased with network depth. This pattern agrees with previous findings in the primate ventral stream (Yamins et al., 2014), where deeper CNN layers better match higher-level visual representations. It therefore suggests that ENTO responses may be more strongly related to higher-level, global, or semantic visual features than to low-level image statistics. In contrast, the ViT model showed a different pattern. Although ViTs did not show the monotonic depth-dependent increase observed in CNNs, their intermediate layers still achieved relatively strong fits (Supplementary Figure S4). One possible explanation is that self-attention allows ViTs to integrate contextual information across the entire image at an earlier stage and thus form global representations earlier than CNNs.

If so, this may indicate that ENTO neurons are sensitive to category-related structure carried by learned global features. At the same time, model architecture itself can influence its similarity to neural responses. For example, recurrent neural networks have been shown to better explain some aspects of IT responses than purely feedforward models (Kubilius et al., 2019). Therefore, further studies are required to clarify the relationship between model architecture and neuronal responses. Driven by advances in deep learning (Bashivan et al., 2019; Jang et al., 2025; Papale et al., 2024; Wang et al., 2024), large-scale collection and analysis of neural responses to natural images is increasingly becoming a major research direction. This shift helps address a major limitation of traditional hand-designed paradigms, which often cannot comprehensively control or sufficiently sample relevant stimulus variables.

On the other hand, an important limitation of the present study is that all neural recordings were obtained under general anesthesia. Although anesthesia is commonly used in avian electrophysiology to improve recording stability, substantial evidence indicates that it can strongly alter cortical/network dynamics, for example, by weakening long-range functional connectivity and effective feedback, changing response gain and temporal profiles, and reducing state-dependent modulation (Suzuki & Larkum, 2020). Notably, our time-resolved analysis (Figure 3A) showed that tuning parameters peaked at 150–170 ms after stimulus onset. This time window overlaps with the late re-entrant feedback dynamics reported in the avian visual system; for example, burst waves in the isthmo-tectal circuit can extend to 150 ms and beyond and may contribute to higher-order integration along the tectofugal pathway (Reynaert et al., 2023). Therefore, the category-related representations observed here may partly emerge during this later integrative stage. Accordingly, the tuning properties and representational geometry reported here are interpreted primarily as stimulus-driven response structures expressed under an anesthetized state, and it is explicitly acknowledged that both the magnitude and temporal dynamics of selectivity may differ in awake, behaving animals. Future studies combining large-scale natural-image or real-world stimulus sets with recordings in awake conditions will be essential for achieving a deeper understanding of the representational mechanisms of ENTO.

DATA AVAILABILITY

The data that support the findings of this study are available in the following repository: https://doi.org/10.6084/m9.figshare.30406729.v1. Source data are provided with this paper.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

J.C.Z., M.J.Z. and Z.Z.W. conceived and designed the study. J.C.Z. and Q.Z.H. performed the surgery and experiments. J.C.Z., X.K.N. and J.T.W. analysed data. P.W. developed the visual stimulus procedure. J.C.Z., M.J.Z. and Z.Z.W. wrote the manuscript. All authors read and approved the final version of the manuscript.

REFERENCES

Anderson C, Parra RS, Chapman H, et al. 2020. Pigeon nidopallium caudolaterale, entopallium, and mesopallium ventrolaterale neural responses during categorisation of Monet and Picasso paintings. *Scientific Reports*, **10**(1): 15971.

Atoji Y, Wild JM. 2012. Afferent and efferent projections of the mesopallium in the pigeon (*Columba livia*). *Journal of Comparative Neurology*, **520**(4): 717–741.

Azizi AH, Pusch R, Koenen C, et al. 2019. Emerging category representation in the visual forebrain hierarchy of pigeons (*Columba livia*). *Behavioural Brain Research*, **356**: 423–434.

Bao PL, She L, McGill M, et al. 2020. A map of object space in primate inferotemporal cortex. *Nature*, **583**(7814): 103–108.

Bashivan P, Kar K, Dicarlo JJ. 2019. Neural population control via deep image synthesis. *Science*, **364**(6439): eaav9436.

Benowitz LI, Karten HJ. 1976. Organization of the tectofugal visual pathway in the pigeon: a retrograde transport study. *Journal of Comparative Neurology*, **167**(4): 503–520.

Bernardi S, Benna MK, Rigotti M, et al. 2020. The geometry of abstraction in the hippocampus and prefrontal cortex. *Cell*, **183**(4): 954–967. e21.

Borg I, Groenen PJF. 2005. Modern Multidimensional Scaling: Theory and Applications. 2nd ed. New York: Springer.

Bushnell BN, Pasupathy A. 2012. Shape encoding consistency across colors in primate V4. *Journal of Neurophysiology*, **108**(5): 1299–1308.

Chang L, Bao PL, Tsao DY. 2017. The representation of colored objects in macaque color patches. *Nature Communications*, **8**(1): 2064.

Clark W, Chilcott M, Azizi A, et al. 2022. Neurons in the pigeon visual network discriminate between faces, scrambled faces, and sine grating images. *Scientific Reports*, **12**(1): 589.

Clark WJ, Colombo M. 2020. The functional architecture, receptive field characteristics, and representation of objects in the visual network of the pigeon brain. *Progress in Neurobiology*, **195**: 101781.

Costalunga G, Kobylkov D, Rosa-Salva O, et al. 2024. Responses in the left and right entopallium are differently affected by light stimulation in embryo. *iScience*, **27**(3): 109268.

Deng J, Dong W, Socher R, et al. 2009. ImageNet: a large-scale hierarchical image database. *In*: Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami: IEEE, 248–255.

DiCarlo JJ, Cox DD. 2007. Untangling invariant object recognition. *Trends in Cognitive Sciences*, **11**(8): 333–341.

Erichsen JT, Hodos W, Evinger C, et al. 1989. Head orientation in pigeons: postural, locomotor and visual determinants. *Brain, Behavior and Evolution*, **33**(5): 268–278.

Farhang E, Toosi R, Karami B, et al. 2025. The impact of spatial frequency on hierarchical category representation in macaque temporal cortex. *Communications Biology*, **8**(1): 801.

Freedman DJ, Riesenhuber M, Poggio T, et al. 2006. Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. *Cerebral Cortex*, **16**(11): 1631–1644.

Freeman J, Simoncelli EP. 2011. Metamers of the ventral stream. *Nature Neuroscience*, **14**(9): 1195–1201.

Griffin, Gregory, Holub A, et al. 2007. Caltech-256 object category dataset. *California Institute of Technology*: (Unpublished), https://resolver.caltech.edu/CaltechAUTHORS:CNS-TR-2007-2001.

Gu Y, Wang Y, Zhang T, et al. 2002. Stimulus size selectivity and receptive field organization of ectostriatal neurons in the pigeon. *Journal of Comparative Physiology A*, **188**(3): 173–178.

Hashimoto W, Kurata K. 2000. Properties of basis functions generated by shift invariant sparse representations of natural images. *Biological Cybernetics*, **83**(2): 111–118.

Hong H, Yamins DLK, Majaj NJ, et al. 2016. Explicit information for category-orthogonal object properties increases along the ventral stream. *Nature Neuroscience*, **19**(4): 613–622.

Hsiao YT, Chen TC, Yu PH, et al. 2020. Connectivity between nidopallium caudolateral and visual pathways in color perception of zebra finches.

Scientific Reports, **10**(1): 19382.

Jang H, Sinha P, Boix X. 2025. Configural processing as an optimized strategy for robust object recognition in neural networks. *Communications Biology*, **8**(1): 386.

Karten HJ, Hodos W. 1968. A stereotaxic atlas of the brain of the pigeon (*Columbia livia*). *The American Journal of Psychology*, **81**(2): 289–290.

Karten HJ, Hodos W. 1970. Telencephalic projections of the nucleus rotundus in the pigeon (*Columba livia*). *Journal of Comparative Neurology*, **140**(1): 35–51.

Kobylkov D, Rosa-Salva O, Zanon M, et al. 2024. Innate face-selectivity in the brain of young domestic chicks. *Proceedings of the National Academy of Sciences of the United States of America*, **121**(40): e2410404121.

Kriegeskorte N, Mur M, Bandettini PA. 2008. Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, **2**: 249.

Krizhevsky A, Sutskever I, Hinton GE. 2017. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, **60**(6): 84–90.

Krützfeldt NOE, Wild JM. 2005. Definition and novel connections of the entopallium in the pigeon (*Columba livia*). *Journal of Comparative Neurology*, **490**(1): 40–56.

Kubilius J, Schrimpf M, Kar K, et al. 2019. Brain-like object recognition with high-performing shallow recurrent ANNs. *In*: Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc. , 1147.

Leutgeb S, Husband S, Riters LV, et al. 1996. Telencephalic afferents to the caudolateral neostriatum of the pigeon. *Brain Research*, **730**(1-2): 173–181.

Levenson RM, Krupinski EA, Navarro VM, et al. 2015. Pigeons (*Columba livia*) as trainable observers of pathology and radiology breast cancer images. *PLoS One*, **10**(11): e0141357.

Li N, Cox DD, Zoccolan D, et al. 2009. What response properties do individual neurons need to underlie position and clutter "invariant" object recognition?. *Journal of Neurophysiology*, **102**(1): 360–376.

Li N, DiCarlo JJ. 2010. Unsupervised natural visual experience rapidly reshapes size-invariant object representation in inferior temporal cortex. *Neuron*, **67**(6): 1062–1075.

Lorenzi E, Kobylkov D, Vallortigara G. 2025. Is there an innate sense of number in the brain?. *Cerebral Cortex*, **35**(2): bhaf004.

Majaj NJ, Hong H, Solomon EA, et al. 2015. Simple learned weighted sums of inferior temporal neuronal firing rates accurately predict human core object recognition performance. *Journal of Neuroscience*, **35**(39): 13402–13418.

Matteucci G, Zoccolan D. 2020. Unsupervised experience with temporal continuity of the visual environment is causally involved in the development of V1 complex cells. *Science Advances*, **6**(22): eaba3742.

Nam Y, Sato T, Uchida G, et al. 2021. View-tuned and view-invariant face encoding in IT cortex is explained by selected natural image fragments. *Scientific Reports*, **11**(1): 7827.

Navarro VM, Wasserman EA, Slomka P. 2020. Taking pigeons to heart: birds proficiently diagnose human cardiac disease. *Learning & Behavior*, **48**(1): 9–21.

Niu XK, Peng YY, Jiang ZY, et al. 2024. Gamma-band-based dynamic functional connectivity in pigeon entopallium during sample presentation in a delayed color matching task. *Cognitive Neurodynamics*, **18**(1): 37–47.

Papale P, De Luca D, Roelfsema PR. 2024. Deep generative networks reveal the tuning of neurons in IT and predict their influence on visual perception. *BioRxiv*, doi: https://doi.org/10.1101/2024.10.09.617382.

Pinto N, Cox DD, DiCarlo JJ. 2008. Why is real-world visual object recognition hard?. *PLoS Computational Biology*, **4**(1): e27.

Pusch R, Packheiser J, Koenen C, et al. 2022. Digital embryos: a novel technical approach to investigate perceptual categorization in pigeons (*Columba livia*) using machine learning. *Animal Cognition*, **25**(4): 793–805.

Reynaert B, Morales C, Mpodozis J, et al. 2023. A blinking focal pattern of re-entrant activity in the avian tectum. *Current Biology*, **33**(1): 1–14. e4.

Rigotti M, Barak O, Warden MR, et al. 2013. The importance of mixed selectivity in complex cognitive tasks. *Nature*, **497**(7451): 585–590.

Soto FA, Wasserman EA. 2016. Promoting rotational-invariance in object recognition despite experience with only a single view. *Behavioural Processes*, **123**: 107–113.

Stacho M, Herold C, Rook N, et al. 2020. A cortex-like canonical circuit in the avian forebrain. *Science*, **369**(6511): eabc5534.

Straight PJ, Gignac PM, Kuenzel WJ. 2024. Mapping the avian visual tectofugal pathway using 3D reconstruction. *Journal of Comparative Neurology*, **532**(2): e25558.

Stringer SM, Perry G, Rolls ET, et al. 2006. Learning invariant object recognition in the visual system with continuous transformations. *Biological Cybernetics*, **94**(2): 128–142.

Suzuki M, Larkum ME. 2020. General anesthesia decouples cortical pyramidal neurons. *Cell*, **180**(4): 666–676. e13.

Wang JT, Rao XP, Huang SM, et al. 2024. Detection of a temporal salient object benefits from visual stimulus-specific adaptation in avian midbrain inhibitory nucleus. *Integrative Zoology*, **19**(2): 288–306.

Wood JN. 2013. Newborn chickens generate invariant object representations at the onset of visual object experience. *Proceedings of the National Academy of Sciences of the United States of America*, **110**(34): 14000–14005.

Wood SMW, Wood JN. 2015. A chicken model for studying the emergence of invariant object recognition. *Frontiers in Neural Circuits*, **9**: 7.

Xiao Q, Li DP, Wang SR. 2006. Looming-sensitive responses and receptive field organization of telencephalic neurons in the pigeon. *Brain Research Bulletin*, **68**(5): 322–328.

Yamins DLK, Hong H, Cadieu CF, et al. 2014. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, **111**(23): 8619–8624.

Yoshida T, Ohki K. 2020. Natural images are reliably represented by sparse and variable populations of neurons in visual cortex. *Nature Communications*, **11**(1): 872.