Title: From Sensory to Perceptual Manifolds: The Twist of Neural Geometry

Running title: Geometric twist of neural manifolds

Authors: Heng Ma[1#], Longsheng Jiang[2#], Tao Liu[1#], Jia Liu[1*]

Affiliation: [1]Department of Psychology & Tsinghua Laboratory of Brain and Intelligence, Tsinghua University, Beijing 100084, China; [2]Beijing Academy of Artificial Intelligence (BAAI), Beijing 100085, China

[#]**Equal contribution**

[*]**Correspondence**: liujiathu@tsinghua.edu.cn

Abstract

To humans, nearly everything is classifiable: whether as big or small, edible or poisonous, righteous or unjust. Similarly, classification is a central task in many machine learning applications, yet the problem of linear inseparability has long posed challenges for artificial neural networks since their inception. Here we asked how biological neural networks tackle this problem by investigating the geometric embedding of neural manifolds in macaques'V2 during orientation discrimination of motion-induced illusory contours. Specifically, we constructed a three-dimensional stimulus space that inherently made the orientation classification of these contours a linearly inseparable problem. As expected, we identified a sensory manifold, formed by the V2 neuron population, that faithfully corresponded to this stimulus space. Crucially, this sensory manifold underwent a series of twist operations, resulting in new axes orthogonal to the original ones. Within this expanded, high-dimensional perceptual manifold, the problem of linear inseparability became linearly separable. Computational models further revealed that the geometric twist operation was achieved by neurons exhibiting nonlinear mixed selectivity in

networks with heterogeneous connectivity patterns. Taken together, our findings elucidate how perception arises from sensation through the lens of neural geometry, enriching our understanding of how cognitive functions derive from underlying anatomical structure.

Keywords: motion-induced illusory contour; nonlinearly inseparable problem; neural geometry; sensory manifold; perceptual manifold; nonlinear mixed selectivity; connectivity structure.

Introduction

Classification constitutes a core cognitive challenge for both biological and artificial intelligence systems, with many tasks potentially reducible to classification problems (Cover, 1965; Chung et al., 2018; DiCarlo et al., 2012; Bernardi et al., 2020). The primary challenge in classification lies in the prevalence of linearly inseparable problems, where perfectly segregating data points into their respective classes using a linear boundary is infeasible (Minsky & Papert, 1969; Fusi et al., 2016). Therefore, machine learning often resorts to complex, nonlinear decision boundaries to address this issue (Schölkopf & Smola, 2002; Williams & Rasmussen, 2006). Here we asked how biological systems address linearly inseparable problems, which may provide insights for developing new machine learning algorithms.

Here we employed an orientation discrimination task involving motion-induced contours (MIC), wherein moving dots generated an illusory contour at the center of a circular viewing area (Fig. 1A). Specifically, dots in one hemisphere of the viewing area moved coherently in a single direction, whereas those in the opposing hemisphere moved contrarily (Chen et al., 2014; Marcar et al., 2000; Mysore et al., 2006). This symmetrical motion led the dots to either

converge or diverge at the center of the viewing area, thus instantaneously forming a left- or right-tilted illusory contour (Fig. 1B). Further, we constructed a three-dimensional stimulus space based on three orthogonal stimulus axes (Lu et al., 2010; Chen et al., 2014; Ma et al., 2021): (1) horizontal versus vertical, (2) outward versus inward, and (3) clockwise or anti-clockwise (Fig. 1C). This resulted in a three-dimensional Euclidean space that embedded the stimulus space within a cylindrical surface (Fig. 1D). Crucially, stimuli sharing the same MIC orientation (e.g., right-titled, red) were interspersed among those with the opposing orientation (i.e., left-tilted, blue), indicating that the classification of MIC orientations presents a linearly inseparable problem within this stimulus space.

Previous neurophysiological studies have revealed that neurons in macaques' V2 show selectivity for cue-invariant MIC orientations (Marcar et al., 2000; Ma et al., 2021), suggesting that the linearly inseparable problem in the stimulus space becomes linearly separable when processed by V2 neurons. Here in this study, we recorded neuronal activity in the V2 to elucidate how the sensory manifold, where MIC orientations remains linearly inseparable akin to the stimulus space, transforms into the perceptual manifold that affords linear separability. To do this, we utilized the population coding doctrine to characterize neural manifolds constituted by neuron population (SueYeon Chung & Abbott, 2021; Seung HS & Lee DD, 2000; DiCarlo & Cox, 2007), an approach recently applied to various domains such as vision (Froudarakis et al., 2020; Dyballa et al., 2023), memory (Xie et al., 2022; She et al., 2021), decision (Okazawa et al., 2021; Flesch et al., 2022; Genkin et al., 2023), navigation (Gardner et al., 2022; Low et al., 2023) and motor execution (Churchland et al., 2012; Russo et al., 2020; Gallego et al., 2017; Gallego et al., 2018). Specifically, we first established the correspondence between the sensory manifold and the stimulus space. Then, we illustrated how twist operations transformed the low-dimensional

sensory manifold into the high-dimensional perceptual manifold by generating new dimensions, where the originally linearly inseparable problem became linearly separable. Furthermore, the expansion of the neural manifold's embedding dimensions required neurons exhibiting nonlinear mixed selectivity in networks with heterogeneous connectivity patterns, thereby echoing the longstanding dictum that structure determines function.

Results

Stimulus space and linearly inseparable problems

Two macaque monkeys were trained to perform an MIC orientation classification task by executing a saccade towards the corresponding side based on the MIC orientation (Fig. 1A). To make the classification a linearly inseparable problem, we designed a three-dimensional stimulus space. For dots in opposing halves, they moved in the opposite directions within a range of $[0, \pi]$ (e.g., dot I and dot II in Fig. 1B, left), and here we used the horizontal direction (i.e., 0) and the vertical direction (i.e., $\pi/2$) as the motion axis of HV (i.e., horizontal versus vertical, Fig. 1C, left). For dots within the same half, one dot (e.g., dot III) either led or trailed another dot (e.g., dot IV) (Fig. 1B, right), quantifiable as the angular location of dot IV with respect to dot III, known as motion shearing. The motion shearing operates within a range of $[0, 2\pi]$ and possesses two independent axes: outward versus inward (OI) and clockwise versus anti-clockwise (CA). Here we chose the shearing directions of $\pi/4$ (inward and clockwise), $3\pi/4$ (inward and anti-clockwise), $5\pi/4$ (outward and anti-clockwise) and $7\pi/4$ (outward and clockwise), which are orthogonally constructed by these two axes (Fig. 1C, right).

Using these three axes of HV, OI, and CA, we constructed a three-dimensional (3-D) Euclidean space within which the stimulus space of a cylindrical surface was embedded, with the

stimuli utilized in this study positioned at the vertices of the inner tangent cube of the cylinder (Fig. 1D). Thus, there were eight global patterns formed by local dot motion, with four patterns exhibiting a right-tilted contour (No. 1 – No.4, red) and the remaining four left-tilted (No. 5 – No.8, blue). Crucially, in the stimulus space, stimuli sharing the same MIC orientation (e.g., right-titled, red) were always interspersed with stimuli exhibiting the opposite orientation (i.e., left-tilted, blue) (Fig. 1D, for the whole set of stimulus parameters, see Fig. S1, also see Movie S1 for the 8 stimuli). In essence, this setup created a classification task that was linearly inseparable. Despite this challenge, both monkeys adeptly performed the task with an accuracy exceeding 90% at 100% motion coherence (Ma et al., 2021).

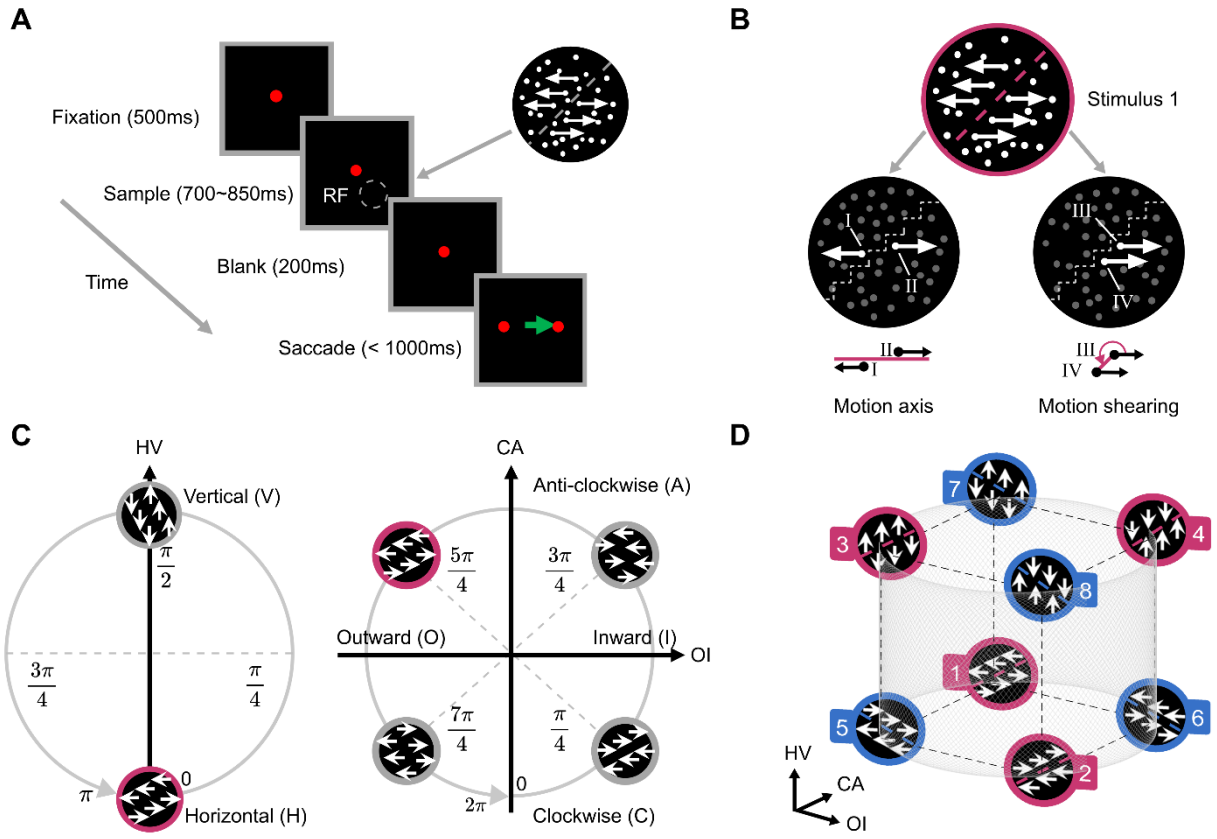

Figure 1. Overview of Stimulus Space.

(A) Left: the MIC orientation discrimination task; right: A right-titled MIC defined by surrounding moving

dots. Monkeys were trained to make saccades according to MIC orientations. This figure is modified from Ma et al. (2021). RF: array population receptive field.

(B) Sample stimulus. The intrinsic dimensions of the stimulus consist of a motion axis (left) and motion shearing (right). The motion axis represents the direction of dots' symmetric movement, for example, dot I and II move horizontally. Right: the motion shearing reveals the relative positioning of adjacent dots moving in the same direction, for example, dot III and IV is $5\pi/4$ in shearing ring, and we define this as outward and anti-clockwise (see panel C).

(C) Axes of stimulus space. Left: the parameters for the motion axis ranging from $[0, \pi]$; Right: the parameters for the motion shearing ranging from $[0, 2\pi]$. From these parameters, three orthogonal axes were extracted: HV (Horizontal vs. Vertical), OI (Outward vs. Inward), and CA (Clockwise vs. Anti-clockwise). Binary values along each axis were used to generate eight stimuli. Within these three axes, the sample stimulus (red) is termed as HOA (Horizontal, Outward, and Anti-clockwise).

(D) Stimulus space. The eight stimuli used in the study, each labeled numerically and color-coded based on their axis locations and MIC orientations (red: right-tilted; blue: left-tilted). Note that the classification of MIC orientations poses a linearly inseparable problem in this stimulus space.

Sensory and Perceptual manifolds

To investigate how the brain addressed this linearly inseparable problem, we recorded 93 neurons in the V2 of the two monkeys to construct a high-dimension neural space (see Methods). First, we investigated how the stimulus space was encoded at the level of neuron population. To do this, we utilized linear support vector machines (SVM) to analyze the time course of the 93 neurons in response to all the 8 stimuli to determine whether the three axes of the stimulus space (Fig. 1D), HV ({1,2,5,6} versus {3,4,7,8}), OI ({1,3,5,7} versus {2,4,6,8}), and CA ({2,3,5,8} versus {1,4,6,7}), could be decoded. We found that classifications along all three axes were linearly separable, with accuracies above 75% and significantly higher than the shuffled baseline

(bootstrap t-tests, all *p*s < 0.001) (Fig. 2A). This finding was replicated with other methods such as the analysis on population average responses and the principal component analysis (Fig. S2). Crucially, the optimal separation direction vectors (determined by SVM, see Methods) for HV, OI, and CA, were mutually orthogonal, as the distribution of the subtended angles between these vectors was significantly more concentrated around 90° than the distribution of angles between two random direction vectors, and also significantly larger than angle distributions between the same classification vectors (Fig. S3).

To visualize the dynamic encoding of the stimuli by the neuronal population, we projected the neural states into a 3-D neural subspace formed by the direction vectors for HV, OI, and CA (see Methods). Figure 2B shows that, prior to stimulus onset, neural states for the stimuli were closely clustered and inseparable. Following the onset, these states gradually spread and became completely separable along the HV axis; then, they remained in a steady phase for the duration of stimulus presentation before returning to their original locations after stimulus offset (see Movie S2 and Fig. S8). Similar dynamic patterns were observed for the OI (Fig. 2C) and CA (Fig. 2D) axes as well. To further illustrate the neural geometry, we utilized the neural states during the steady phase, ranging from 200ms to 500ms post-stimulus onset. Fig. 2E shows that each stimulus-associated neural state was located at a vertex of a slightly distorted cube. Crucially, the geometric relation among these neural states corresponded to the geometric relation among stimuli in the original stimulus space (Fig. 1D); accordingly, we refer to the neural states embedded in this 3-D subspace as the “sensory manifold”. Since the function of sensation is to convert environmental energy into neural signals, the sensory manifold was sensitive to variations in the coherence of moving dots in the stimuli, which gave rise to a series of concentric cubic sensory manifolds with higher coherence levels eliciting larger cubic manifolds

(Fig. S5). Taken together, the sensory manifold faithfully represented the stimulus space, with HV, OI, and CA serving as the axes of the neural subspace that embedded this manifold. Consequently, the linearly inseparable problem cannot be solved within the sensory manifold.

Consistent with previous studies on individual neurons (Marcar et al., 2000; Ma et al., 2021), our SVM analysis at the population level was able to linearly decode MIC orientations (i.e., left- versus right-titled, LR) (Fig. 2A, orange line), and the accuracy was comparable to that of decoding the stimuli's features (i.e., HV, OI, and CA). Fig. 2F illustrates the neural states along the optimal separation vector for LR, where the linearly inseparable problem became linearly separable along this axis (for dynamic trajectories see Fig. S8 & Movie S2). Critically, this LR axis was orthogonal to the three existing axes for the sensory manifold (Fig. S3A), suggesting that the LR axis emerged not directly from the stimuli but from the processing of sensory information. Evidence supporting this conjecture comes from the analysis on the latency of the emergence of the LR axis, which occurred about 30ms after the emergence of the axes encoding the stimuli's features (i.e., HV, OI, and CA) (Fig. 3F). This observation is line with previous findings on visual motion segregation (Braddick, 1993; Lamme, 1995; Chen et al., 2014; Hu et al., 2018), echoing the idea that perception is a result of operations on sensory information (e.g., Goldstein, E. B., & Cacciamani, L, 2021). Accordingly, we refer to the manifold embedded in this dimension-expanded neural subspace, which now incorporated the new LR axis in addition to the existing axes for the sensory manifold, as the "perceptual manifold". Next, we explored the potential operation that transforms the sensory manifold into the perceptual manifold.

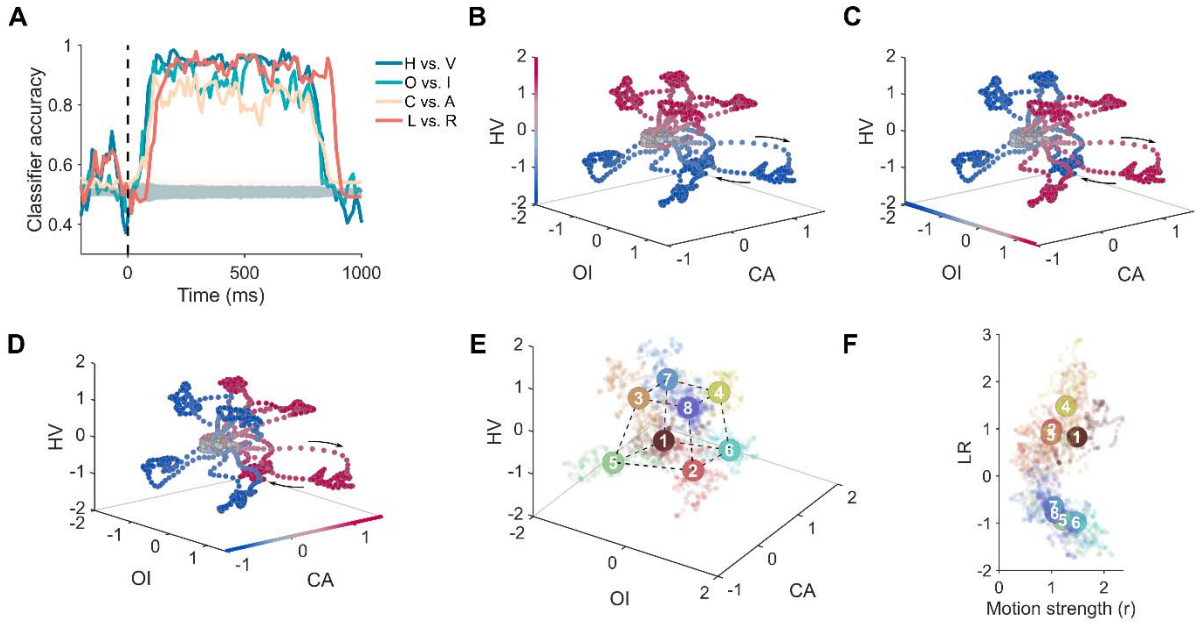

Figure 2. Neural geometry of sensory and perceptual manifolds.

(A) Time-course accuracies by linear SVM classifiers. The SVMs were trained on neural data from 93 neurons within the time interval [-200ms, 1000ms] relative to stimulus onset. Shaded regions denote the mean SVM accuracy (±3 standard deviations) of the baselines by shuffling stimulus labels 1000 times. Individual monkey's results are shown in Fig. S6 E&F.

(B-D) Dynamic trajectories of neural states. Neural states change from -50ms to 1000ms were projected onto the neural subspace defined by the HV, OI, and CA axes. Black arrows indicate the direction of neural state movement. Blue and red colors indicate categorization along the HV, OI, and CA axes for (B), (C), and (D), respectively. Individual monkey's results are shown in Fig. S7.

(E) Sensory manifold. Eight clusters of neural states in the steady phase (200ms-500ms after stimuli onset) corresponded to the 8 stimuli. The centers of these clusters are connected with dashed lines. Cool and warm colors indicate the two MIC orientations, respectively. Individual monkey's results are in Fig. S6 A&B.

(F) Dynamic trajectories of neural states along the LR axis. The clusters follow the same color code as in panel E, with the centers of the clusters marked with solid circles. Note that cool and warm colors are separated along the LR axis. Motion strength ( ), which is the Euclidean distance between the neural states and the origin in the sensory manifold in panel E, is included for display purposes only.

Twist operations on neural geometry

For simplicity, here we use x, y, and z to denote the axes of HV, OI, and CA, respectively. In the z-x plane, the four vertices $\{1,3,5,7\}$ exemplify the standard planar XOR problem (Fig. 3A, left). Traditional solutions to the XOR problem often involve expanding dimensionality (Minsky & Papert, 1969; Fusi et al., 2016). For example, twisting two parallel edges along the z-axis in opposite directions around the x-axis can resolve the issue (Fig. 3A, middle), thus termed as twist operation (for a similar idea, see Nogueira et al., 2023). Accordingly, this twist operation transforms the 2-D plane into a curved 3-D surface (Fig. 3A, right). In this new 3-D space, a new axis, denoted as v, is constructed through a cross-product-like operation between z and x, expressed as $v = z \oplus x$. Along this new v axis, the four vertices become linearly separable.

To examine whether this theoretically derived v axis was present in the actual neural space, we trained linear SVMs to classify three sets of vertex pairs among four vertices. As expected, the vertex pair of $\{1,5\}$ versus $\{3,7\}$ and that of $\{1,7\}$ versus $\{3,5\}$ were linearly separable along the x axis and z axis, respectively, with accuracies greater than 75%. Critically, the vertex pair of $\{1,3\}$ versus $\{5,7\}$ was also linearly separable along the v axis, suggesting that the v axis was indeed present in the neural space (Fig. 3B). Next, we generalize this twist operation from the planar XOR problem to the cubic XOR problem (i.e., {1, 2, 3, 4} versus {5, 6, 7, 8}, Fig. 3C, left).

Apparently, a single twist operation with $v = z \oplus x$ proves insufficient to resolve the cubic XOR problem (Fig 3C, middle). However, incorporating a second twist in a plane parallel to the v axis around the y axis makes the vertex pair of {1, 2, 3, 4} versus {5, 6, 7, 8} linearly separable (Fig 3C, right). In this second twist, a new axis, denoted as p, is constructed through a cross-product-like operation between v and y: $p = v \oplus y = z \oplus x \oplus y$. Accordingly, this

model is thus termed as the double-twist model. Interestingly, this newly constructed axis is the axis that differentiates between the left- and right-titled MIC orientations. Indeed, the axis is the LR axis illustrated in Fig. 2F.

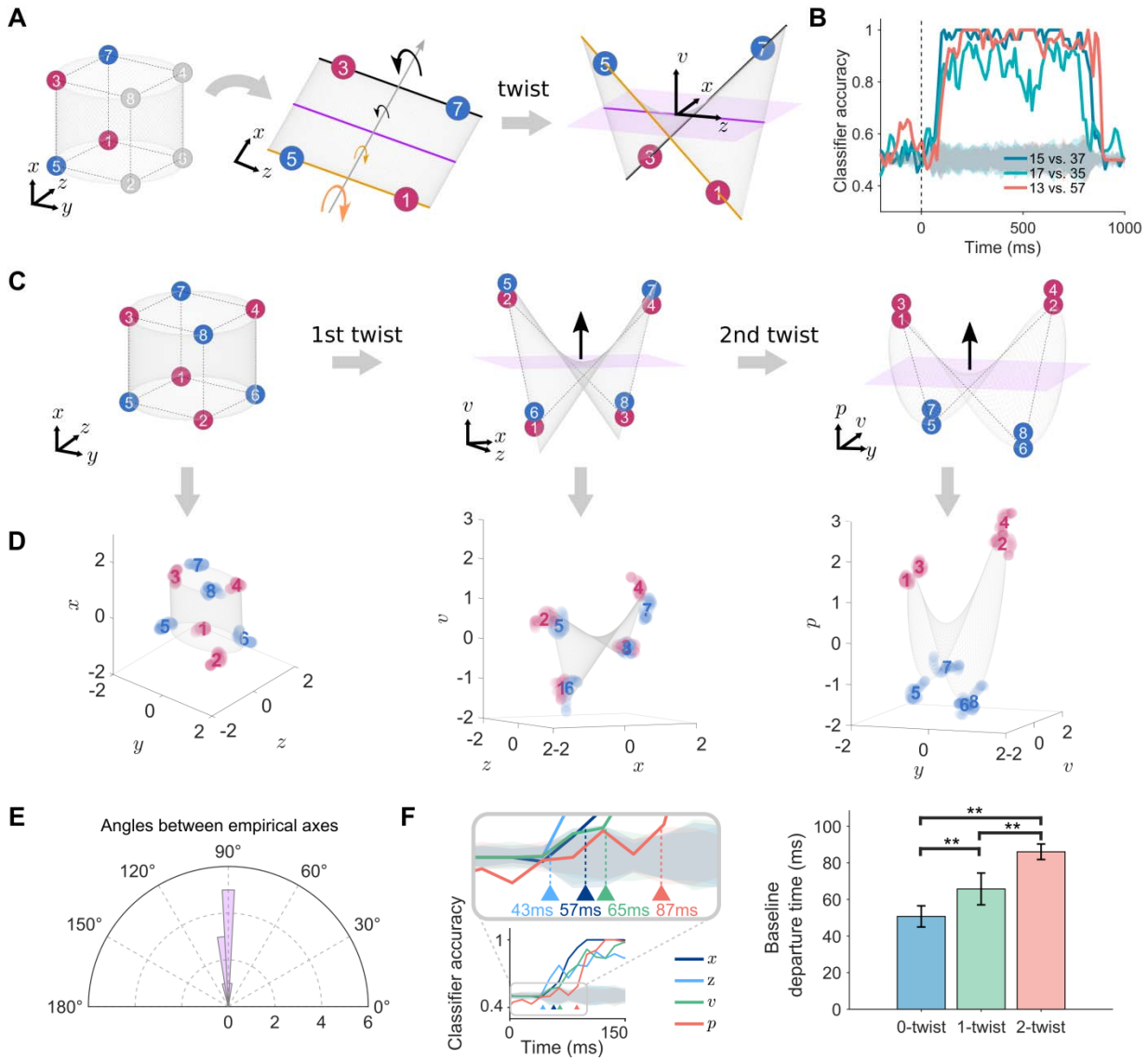

Figure 3: The double-twist model.

(A) One twist operation on a planar XOR problem. Left: the coplanar vertices are colored based on MIC orientations, forming a standard XOR problem on the - plane. Middle: a twist operation along the z-axis resolves the XOR problem by rotating two parallel edges in opposite directions around the x-axis. Right: this twist operation generates the axis orthogonal to the - plane.

(B) Time courses of SVM classification accuracies. Shaded regions denote the mean SVM accuracy (±3 standard deviations) of the baselines by shuffling stimulus labels 1000 times.

(C) Double twist operations on a cubic XOR problem. Two consecutive twists are required to correctly classify the 8 vertices according to their MIC orientations. Left: the theoretical sensory manifold in the y-z-x space; Middle: the intermediate manifold achieved after one twist operation on the z-x plane, with the emergence of the new v axis. Right: the task-relevant p axis emerges after the second twist operation on the y-v plane.

(D) Neural manifolds corresponding to the theoretical manifolds achieved through twist operations. The axes of x, y, z, v and p in the neural space were achieved through corresponding SVM classifiers. Eight clusters of neural states corresponded to the 8 stimuli, and the centers of these clusters were used to depict the geometry of neural manifolds through linear fitting.

(E) Angle distribution between x, y, z, v and p axes. The distribution of the subtended angles between the optimal separation direction vectors as determined by SVM was concentrated around 90°.

(F) Latency of axis emergence. Left: time courses of classification accuracies of the axes. The enlarged insert shows the latencies of the emergence of the z, x, v, and p axes (against their corresponding baselines) at 43ms, 57ms, 65ms, and 87ms, respectively. Shaded regions denote the mean SVM accuracy (±3 standard deviations) of the baselines by shuffling stimulus labels 1000 times. Right: The latencies of the sensory axes (0-twist, blue), the intermediate axes (1-twist, green), and the task-relevant axis (2-twist, red). **: $p < 0.001$.

To visualize the perceptual manifold embedded in the neural subspace constructed by the z, x, y, v, and p axes, we projected the neural states during the steady phase (i.e., 300ms–500ms) onto a 3-D subspace constructed by three axes selected. Figure 3D shows the geometric transformation of the sensory manifold (left) into the perceptual manifold (right), which are equivalent to two successive twist operations (see also S7A). After the first twist operation, two planar XOR problems were resolved: the vertex pair of {1, 3} versus {5, 7} and that of {2, 4} versus {6, 8} became linearly separable along the intermediate v axis (Fig. 3D, middle). After the second twist, the cubic XOR problem was resolved along the task-relevant p axis, and the

perceptual manifold emerged accordingly (Fig. 3D, right). In addition, the v and p axis, as well as the z, x, and y axes, were mutually orthogonal (Fig. 3E). Note that the neural manifolds derived from actual neural states (Fig 3D) closely matched the theoretical manifolds generated through twist operations (Fig 3C) (all $R^2 \geq 0.8$, for details see Methods), suggesting that the neural manifolds observed in the macaques' V2 may indeed undergo geometric transformations equivalent to those theorized twist operations.

More direct evidence supporting this conjecture comes from the analysis on the latency of the emergence of the intermediate v axis. Our double-twist model predicts that the v axis should emerge after the sensory axes of z and x, upon which the first twist operation acts, and before the task-relevant p axis, which relies on the v axis for the second twist. Consistent with this prediction, the v axis emerged (against the baseline) at 65ms post-stimulus onset, significantly earlier than the emergence of the z and x axes (at 43ms and 57ms respectively; bootstrap t-test, $p < 0.001$), yet later than that of the p axis (at 87ms; bootstrap t-test, $p < 0.001$) (Fig. 3F, left; for the time courses of all intermediate axes, see Fig S7 B & C).

Taken together, the presence of the MIC-unrelated v axis, along with the observed temporal sequence among the different axes, suggests that the perceptual manifold did not directly derive from physical stimuli; instead, it appears to transition through a sequence of intermediate transformations from the sensory manifold, each adhering to essential geometric procedures required for evolving from one manifold to another. Next, we explored the possible existence of other intermediate axes similar to the v axis.

Dimensionality of perceptual manifold

Given the commutative nature of the twist operation, $p = z \oplus x \oplus y$ can alternatively be expressed as $p = x \oplus y \oplus z$ or $p = y \oplus z \oplus x$. Thus, our double-twist model predicts the existence of two additional intermediate axes: $u = x \oplus y$, which classifies vertices $\{1,4,5,8\}$ from vertices $\{2,3,6,7\}$, and $w = y \oplus z$, which classifies vertices $\{1,2,7,8\}$ from vertices $\{3,4,5,6\}$. Consistent with this prediction, the SVM analysis showed significantly higher classification accuracies for these vertex pairs compared to the baseline, confirming the existence of these two intermediate u and w axes in the neural space (Fig. S10A). The characteristics of the u and w axes were the same as those of the v axis: first, the u and w axes were orthogonal to each other and to the other axes (Fig. S10B); second, they emerged subsequent to the sensory axes (i.e., z, x, and y axes) yet prior to the p axis ($F(2, 697) = 1015.6$, $p < 0.001$; for all pair-wise comparisons, $ps < 0.001$, Bonferroni corrected) (Fig. 3F, right; also see Fig. S10C); third, the cubic XOR problem can be addressed using either the u axis or the w axis as the intermediate axis (Fig. S9). In sum, seven axes, the x, y, z, u, v, w and p axes, were identified in the neural space of the neuron population in the V2, suggesting that the perceptual manifold was embedded in a 7-D neural subspace, a dimensionality much higher than that of the sensory manifold (i.e., 3-D).

An intriguing question arises: why was the perceptual manifold embedded in a 7-D space when a 4-D space, constructed by x, y, z, and p axes, would suffice for the LR classification in the MIC task? One possibility is that the availability of multiple alternative pathways to construct the task-relevant p axis enhances the robustness for the LR classification. Alternatively, the perceptual manifold may not be task-specific; rather, the LR classification could be just one of its many possible applications. That is, for the stimulus space with 8 vertices, there are 2^8=256 possible classifications. Some of these are linearly separable, like vertex $\{1\}$ versus

$\{2,3,4,5,6,7,8\}$ or vertices $\{1,5\}$ versus $\{2,3,4,6,7,8\}$ (Fig. 4A), while others are not, such as $\{1,3\}$ versus $\{2,4,5,6,7,8\}$ (Fig. 4B). Let a linear classification be X versus X^c with $\|X\| = 0, \ldots, 8$, there exist 104 different types of linear classifications, leaving the rest 152 classifications are linearly inseparable (for a full list, see Fig. S11 and Table S1). Critically, these linearly inseparable classifications in the stimulus space become linearly separable in the 7-D space. For example, the vertex pair of $\{1,3\}$ versus $\{2,4,5,6,7,8\}$ becomes linearly separable in the y-z-p subspace (Fig. 4C). In total, there are $C_3^7 = 35$ 3-D subspaces (Fig. S12), and each of the 152 linearly inseparable classifications becomes linearly separable in at least one of these 35 subspaces (Table S2). That is, within the 7-D space, every conceivable classification in the stimulus space becomes linearly separable.

Therefore, the perceptual manifold is not the manifold for decision-making or action; rather, it provides a reservoir of all possible candidates (in our case, solutions for all conceivable classifications), awaiting selection by higher-order cognitive processes. In support of this idea, we found that, upon the removal of neurons highly sensitive to MIC orientations (i.e., task-relevant neurons) from the V2 neuron population, the remaining neurons with mixed selectivity were still capable of decoding MIC orientations at the population level, just as effectively as the original neuron population (Fig. S4). This finding suggests that neurons with mixed selectivity to various features (Rigotti et al., 2013; Kaufman et al., 2022) may play a critical role in transforming the sensory manifold into the perceptual manifold.

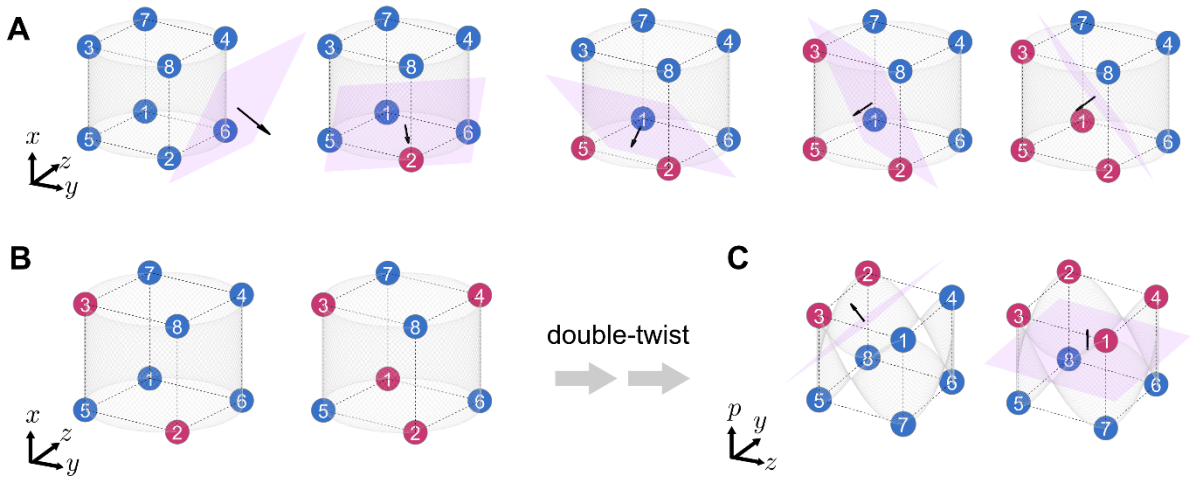

Figure 4: Linearly separable and inseparable classifications in the stimulus space.

(A) Examples of linearly separable but non-axial classifications in the stimulus space. The classification separates a subset of vertices from the rest with 0, 1, 2, 3, and 4 (left to right, red vertices). The classifications with 5, 6, 7, and 8 are complementary (blue vertices). In total, there are 104 linearly separable classifications in the stimulus space when the axial classifications are included.

(B) Examples of linearly inseparable classifications in the stimulus space, including the LR classification (right). In total, there are 152 linearly inseparable classifications in the stimulus space.

(C) The classifications in panel B become linearly separable in new subspaces obtained through twist operations (see Table S1).

Nonlinear mixed selectivity neurons and their heterogeneous connectivity patterns

To exemplify the role of neurons with different response profiles in expanding dimensionality of neural manifolds, we examined a one-layer network consisting of three neurons, , and . These neurons receive inputs consisting of two continuous features, and , that form a 2-D feature sheet (Fig. 5A, top). The XOR problem, in this context, is defined as categorizing points located at diagonally opposite corners (i.e., red versus blue points). The neurons receive either pure (either or) or mixed (the combination of and) inputs, and possess either linear or

nonlinear (e.g., ReLU) activation functions, leading to four distinct output patterns (Fig. 5A, bottom).

Neurons with pure selectivity and the linear activation function (i.e., linear pure selectivity) afford only a simple affine transformation of the original feature sheet (Fig. 5B). Because the rotation axis is aligned with the orange and black edges, the edges remain parallel after transformation. For neurons with nonlinear pure selectivity, the feature sheet undergoes bending because of the nonlinear activation function (Fig. 5C). Because the bending occurs along an axis parallel to the orange and black edges, the edges remain parallel after bending. For neurons with linear mixed selectivity (Fig. 5D), the feature sheet endures a complex affine rotation, with the rotation axis not aligning with the three main axes. However, even after this rotation, the edges remain parallel. In these three scenarios, neither bending nor rotation alone suffices to expand the dimensionality of the feature sheet, and thus the red and blue points remain linearly inseparable.

In contrast, neurons with nonlinear mixed selectivity (NMS) perform a twist operation on the feature sheet, which derives from the combination of affine rotation and the bending. After the twist operation, the orange and black edges are no longer parallel (Fig. 5E), making the XOR problem linearly separable. In sum, this theoretical analysis suggests that NMS neurons likely play a critical role in addressing the XOR problem.

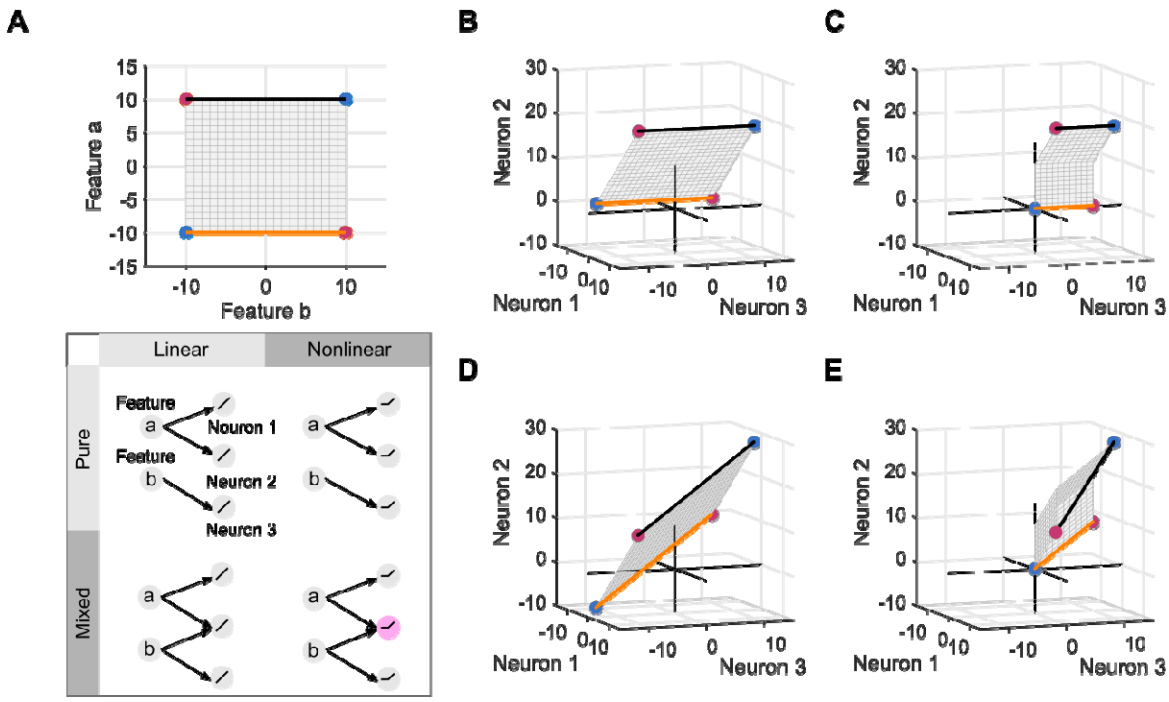

Figure 5: Neurons with different response profiles in realizing twist operations.

(A) Top: The 2-D feature sheet represents all possible combinations of stimulus features and . Two edges of this sheet are highlighted as orange and black, and four corners are colored red and blue to indicate different categories. Bottom: Four types of neurons' response profiles derive from the combination of input (pure versus mixed) with activation function (linear versus nonlinear) activation functions. ReLU is used for the illustration of nonlinear activation functions. The neuron marked in light purple denotes the NMS neuron.

(B-E) The representations in the neural space constructed by (B) three linear pure selectivity neurons, (C) three nonlinear pure selectivity neurons, (D) two linear pure selectivity neurons and one linear mixed selectivity neuron, or (E) two nonlinear pure selectivity neurons and one NMS neuron.

Further formal proof shows that the unique characteristic of NMS neurons lies in their interactive responsiveness to stimulus features (see Methods for details). For example, the time course and average activity of a typical V2 neuron shows that this neuron responded differently to HV and CA while showing no sensitivity to OI (Fig. 6 A & B, top). Thus, we used the

significance of a three-way interaction among the stimulus features (HV, OI and CA) as an index to quantify whether a neuron exhibits nonlinear mixed selectivity (see Methods). We found that all V2 neurons recorded showed significant interactions (with a minimum ,), indicating that they were all NMS neurons. To simulate neurons with pure selectivity (Rigotti et al., 2013), we generated synthetic neurons exclusively tuned to one of the three features (Fig. S15). For example, Fig. 6 A & B (bottom) shows a typical synthetic neuron with pure selectivity to OI. Next, we examined the performance of both NMS and pure selectivity neurons in addressing problems that are either linearly separable or inseparable.

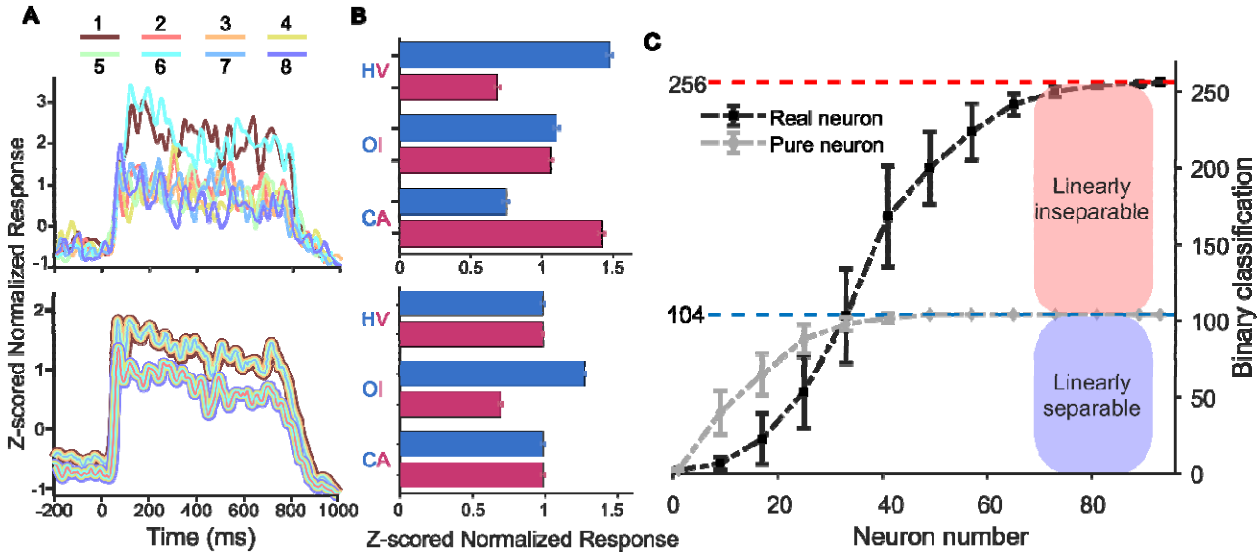

Figure 6: Classification performance of NMS and pure selectivity neurons.

(A) Time courses of an NMS neuron (top) and a synthetic neuron with pure selectivity (bottom). Colors and numbers denote different stimuli and their associated neural responses. The time course of the synthetic neuron was generated by modifying the time course of the NMS neuron shown above.

(B) Average activity of the NMS neuron (top) and the synthetic neuron (bottom). Activity magnitude represents the average activity from 200ms to 750ms after stimulus onset. Colors denote different conditions of a feature, and error bars denote the standard error of mean.

(C) The number of successful linear classifications increases as a function of neuron population size for both pure selectivity neurons (gray) and NMS neurons (dark). The gray curve plateaus at 104, the total number of linearly separable problems in the stimulus space, whereas the dark curve converges to 256, which includes 152 linearly inseparable problems in addition to 104 linear separable ones. Real neuron: neurons recorded in the V2; Pure neuron: synthetic neuron with pure selectivity; error bar: standard deviation.

To evaluate the classification efficacy for 152 linearly inseparable problems (Table S1) and 104 linear separable problems (Table S2) in the stimulus space, we used a population-increment procedure, where the population size was progressively increased from a single neuron to the entire set of 93 neurons, adding one randomly selected neuron each time. During each iteration, we trained SVMs with neural activities for all possible classifications, and a classification accuracy threshold of 75% was set as the criterion for successful classification (for details, see Methods). Fig. 6C shows the number of successful linear classifications (i.e., binary classification) as a function of neuron population size. With NMS neurons, we achieved the full range of possible classifications (256 in total) for both linearly separable and inseparable problems (Fig. 4C, black curve) once the number of NMS neurons exceeded 81. In contrast, using the synthetic neurons with pure selectivity, the total number of successful classifications plateaued at 104 (Fig. 4C, gray curve) once the population size exceeded 49. That is, additional increases in population size did not improve classification performance. Critically, close examination of the problems successfully classified by pure selectivity neurons revealed that they were all linearly separable in the stimulus space (Table S1), and none came from the set of linear inseparable problems (Table S2). That is, regardless of population size, neurons with pure selectivity can only address linearly separable problems. Taken together, our finding suggests that pure selectivity neurons alone cannot expand the sensory manifold to a higher

dimensionality; instead, NMS neurons are necessary in expanding the dimensionality of neural manifolds, hereby transforming the sensory manifold into the perceptual manifold.

While necessary, NMS neurons alone seem insufficient for dimension expansion, as it has been suggested that the dimensionality of neural manifolds is also dictated by the structure of neural networks (Langdon et al., 2023). Here we further investigated the connectivity structure of NMS neurons that may underpin dimension expansion. To easily quantify the connectivity structure, we built a two-layer feedforward neural network tasked with processing the stimuli used in the macaques' experiment (for details on the network, see Methods). In this network, each output neuron receives the combination of all three stimulus features from the input neurons (i.e., HV, OI and CA) and employs a nonlinear activation function, with connectivity weights independently sampled from a Gaussian distribution. As a result, all output neurons in the neural network demonstrate a response profile of nonlinear mixed selectivity.

In this network, the connectivity structure is controlled by parameter d, which signifies the degree of diversity in connection patterns among the two layers (Fig. 7A). This diversity ranges from identical patterns ($d = 0$) to completely uncorrelated patterns ($d = 1$) (see Methods). That is, when d = 1, each NMS neuron generates a distinct response because the connection pattern from the input neurons is unique (Fig. 7A, right). Therefore, the matrix of connectivity weights is full rank, and after passing through the nonlinear activation function this connectivity structure allows the neural manifold to carry out all possible linear and nonlinear classifications composed by the stimulus features. Indeed, the SVM analysis, similar to that on the macaques' data, was performed, and a 7-D perceptual manifold was identified. For visualization, neurons' activations are projected into 3-D subspaces (Fig. 7B), where each dot denotes the neural state of a stimulus, with red and blue colors representing the two MIC orientations. Within this 7-D perceptual

manifold, we can identify the sensory manifold embedded in a 3-D subspace constructed by axes corresponding to the three stimulus features (Fig. 7B, left), the intermediate manifold in a 3-D subspace with a new axis of $v = z \oplus x$ after one twist operation (Fig. 7B, middle), and the subspace achieved after the second twist operation where linearly separating MIC orientations becomes possible (Fig. 7B, right). In addition, we also used continuous stimuli that spanned the entire motion-shearing rings (Fig. 1C, right) and found similar results (Fig. S16). In sum, this 7-D perceptual manifold constructed by the network of NMS neurons with a random connectivity (i.e., $d = 1$) is similar to the 7-D perceptual space identified in the macaque's V2 (Fig. 3D).

In contrast, when $d = 0$, all NMS neurons have the same inputs and thus generate identical responses, similar to neurons in a theoretical functional module that have the same response profile. Accordingly, the matrix of connectivity weights of the network is rank 1 (or 0 if all weights are 0), and thus the dimensionality of the neural manifold generated by this network is low (dimensionality = 1.3, std = 0.70). This can be clearly demonstrated with PCA analysis on the variance of neurons' activation of the network (Fig. 7D). When $d = 0$, the first PC explained 99.84% of total variance, leaving nearly no variance for the rest of PCs. Thus, the neural states of the stimuli were confined into an approximately 1-D space (Fig. 7C). In contrast, when $d = 1$, to explain the same amount of variance of first PC when $d = 0$, the first 6 PCs (99.78%) were required. In other words, when $d = 1$, the neural states of the stimuli were dispersed into a higher dimensional space. Therefore, unlike the network with high diversity in connectivity (i.e., d = 1), the neural manifold constructed by the network with no diversity (i.e., d = 0) suffers a severe deficit in performing either linear or nonlinear classification (number of linearly separable problems successful addressed = 9.16 or 8.8% of the whole set, std = 2.89; number of linearly

inseparable problems successful addressed = 2.74 or 1.8%, std = 2.23). That is, networks with identical connection patterns hardly encode much information, even with all neurons showed a response profile of nonlinear mixed selectivity.

To systematically investigate how connectivity structure influences neural manifolds, we constructed a series of neural networks with different parameter d, and then measured the dimensionality of the neural manifolds and the classification performances on both linearly separable and inseparable problems (see Methods). We found that as the diversity of connectivity patterns increased, the dimensionality of the neural manifolds also increased monotonically (Fig. 7E, red curve). Interestingly, the network did not need to possess a completely diverse structure to support a high-dimensional neural manifold. With $d > 0.5$, the dimensionality was reliably expanded to 7. Accordingly, the number of successful classifications increased monotonically, finally capable of successfully classifying all possible problems (i.e., 256) after the dimensionality reliably reached 7 (Fig. 7E, blue curve). Note that the linearly inseparable problems were resolved in parallel with the linearly separable classifications (Fig. S17).

In sum, NMS neurons alone are not sufficient; the heterogeneity of NMS neurons derived from heterogenous connectivity patterns is also critical, which optimally leverages the NMS neurons to construct a higher-dimensional neural space.

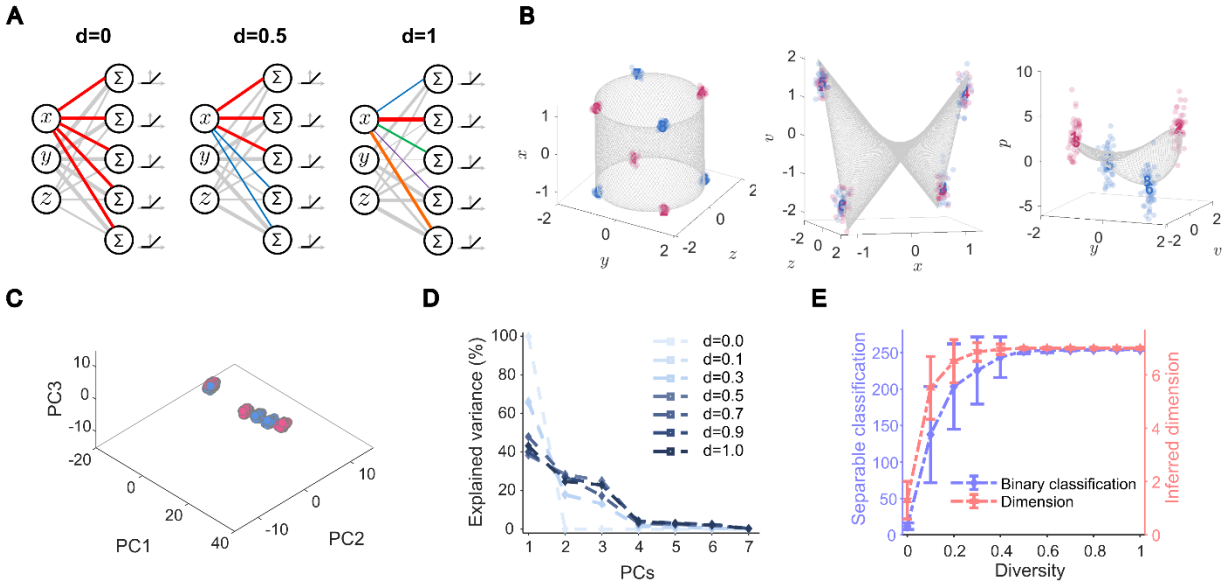

Figure 7. Connectivity structure on dimensionality.

(A) Schematic illustration of a two-layer feedforward neural network. Each output neuron of the network receives the inputs encoding three stimulus features (HV, OI and CA), indexed by , , and , respectively, from the input neurons. The connectivity weights are randomly sampled from a multivariate Gaussian distribution, with the thickness of lines indicting the magnitudes of weights. The connectivity pattern from feature to the output neurons is highlighted by color for display purpose. Output neurons' response is achieved through linear summation and ReLU activation. Networks with three levels of diversities in connectivity patterns are shown here.

(B) Visualization of neural manifolds in 3-D subspaces when = 1. Left: the sensory manifold that corresponds to the stimulus space; Middle: the intermediate manifold after one twist operation ( ); Right: the subspace where linearly separating MIC orientations is achieved. Each point denotes the neural state of a stimulus, with red and blue colors representing the two MIC orientations (LR), respectively.

(C) Visualization of neural manifold in a 3-D subspace when = 0. The subspace is constructed by the first three PCs of neurons' activation.

(D) Variance explained by PCs of neurons' activation with different connectivity structures. When = 0, indexed by the dashed line with the lightest blue, the variance is nearly explained by the first PC alone. When

$d = 1$, indexed by the dashed line with the darkest blue, the curve of variance explained by each PC is flattened.

(E) The dimensionality and the number of successful classifications as the function of connectivity heterogeneity (diversity) of networks. Left axis (blue): the number of successful binary classifications. Right axis (red): the effective dimensionality of the neural manifolds constructed by the network. Error bar: standard deviation.

Discussion

Sensation is typically defined as the conversion of environmental energy into neural signals, which are subsequently interpreted and give rise to perception. However, the lack of conceptual clarity often leads to debates on whether the dichotomy between sensation and perception is real and necessary (e.g., Gibson, 1979), and if real, where and why the line of demarcation shall be drawn. Here we investigate how physical moving dots acting upon sense organs are interpreted as an illusory contour through the lens of the population doctrine. By analyzing neural geometry constructed by neuron population in macaques' V2, we found that the stimuli were faithfully represented in a 3-D sensory manifold. Critically, to linearly decode the orientation of the illusory contours, an inherently linearly inseparable problem in the stimulus space, we identified four additional axes, along with three axes for the sensory manifold, that constructed a 7-D perceptual manifold through geometric transformations equivalent to twist operations. Critically, the geometric transformation from the sensory manifold to the perceptual manifold was realized through collaborative work combining specialized neuronal properties of nonlinear mixed selectivity at neuron level with heterogeneity in response profiles at population level. In summary, our study demonstrated an unambiguous line drawn between sensation and perception

from a neural geometry perspective, enriching our understanding to the dictum that structure determines function in neural networks (Langdon et al., 2023; Sporns, 2014).

Previous studies on neural geometry have shown that neural manifolds can faithfully represent both stimulus space (Bao et al., 2020; Xie et al., 2022; Gardner et al., 2022) and action space (Churchland et al., 2012; Perich et al., 2020; Gallego et al., 2017). In line with these findings, our study revealed a cubic-shaped neural manifold encoding three mutually orthogonal axes, each of which corresponded to HV, OI, and CA feature of stimuli, respectively. In addition, the manifold's size was monotonically increased as a function of the motion coherence of local dots. That is, this manifold, generated by stimuli acting upon the eyes, is therefore termed the sensory manifold. Further, we identified four additional axes encoding features absent from the physical stimuli. Specifically, one such axis, geometrically equivalent to the double twist operations on three stimulus features, encoded perceived orientations of the illusory contours. Consequently, the neural geometry transformed from the 3-D sensory manifold into the 7-D perceptual manifold within a delay of about 30ms. This analysis may provide a formal account of the dynamics involved in the hierarchical abstraction of categories (Dehaqani et al., 2016). Further, the perceptual manifold was task irrelevant, because in addition to differentiating the MIC orientations, it is equipped to accomplish the remaining 255 possible classifications, irrespective of linear separability. Therefore, the task irrelevance characteristics of the perceptual manifold also draws a line of demarcation between the perceptual manifold and manifolds associated with decision-making or action.

The creation of these perception-related axes was solely attributed to NMS neurons. On one hand, the populational responses of NMS neurons are able to change the geometry of neural manifolds equivalent to twist operations, thus expanding dimensionality of neural manifolds

(Rigotti et al., 2013; Fusi et al., 2016). On the other hand, neurons either purely selective for a single feature or exhibiting a linear mix of multiple features selectivity showed their capability only to address linearly separable problems, suggesting the necessity of NMS neurons in producing subjective experiences of perception. Unfortunately, the functionality of these neurons was largely ignored under the neuron doctrine, where attention was predominantly directed to neurons with clear semantic meanings (i.e., neurons with pure selectivity). In contrast, our study advocates the indispensable role of NMS neurons in geometric transformations from the sensory manifold to the perceptual manifold and generating latent variables from intermediate axes that did not have semantic description (Jazayeri & Ostojic, 2021). Therefore, NMS neurons shall lie at the core of the population doctrine.

However, the mere presence of NMS neurons is not sufficient, and their functional efficacy is intrinsically tied to the heterogeneity at population level (Kriegeskorte & Wei, 2021; Langdon et al., 2023). With the simulation of networks with NMS neurons, we found that when the connectivity patterns to NMS neurons were in a homogenous fashion, the expansion of neural dimensionality was limited, thus constraining the network's computational power in classifying either linearly separable or inseparable problems. Conversely, when the heterogeneity in the connectivity patterns was increased, a significant expansion in the dimensionality of neural manifolds was observed, effectively transforming linearly inseparable problems into linearly separable ones. Taken together, our study revealed the symbiotic relationship between the response properties of nonlinear mixed selectivity at neuron level and the heterogeneity of NMS neurons at population level. In this framework, NMS neurons serves the "building blocks" that enable the construction of high-dimensional neural manifolds, whereas the specific architectural configuration among them provides the "blueprint" to actualizes this potential. That is, both

specialized neuronal properties and the intricate configuration are critical in sculpting the landscape of neural computation.

The high dimensionality of the perceptual manifold provides a reservoir of all computational solutions, enabling flexible classifications according to downstream task rules. However, this flexibility comes with a cost as the potential solutions rise exponentially with the number of dimensions. For example, in our study, downstream cortical regions must identify one proper classification required by the task from 256 possibilities. While the mechanism for effectively searching solutions based on task demands remains largely unknown, the strengthening or weakening of the homogeneity of connectivity patterns through Hebbian ("fire together wire together") and anti-Hebbian ("out of sync, lose the link") rules might hint at a possible mechanism of reducing dimensionality (Caporale & Dan, 2008). In networks with Hebbian plasticity, neurons that are frequently co-activated by similar tasks or stimuli often develop stronger and more homogeneous connectivity patterns, and thus increase the selectivity to specific features and lead to the formation of specialized modules. Within these modules, neurons tend to exhibit heightened pure selectivity, thus reduced mixed selectivity, at neuron level and decreased heterogeneity at population level (Weigand et al., 2017; Lindsay et al., 2017; Bullinaria, 2007). On the other hand, with the anti-Hebbian rule, task-relevant features are disentangled from task-irrelevant ones. Accordingly, axes associated with less pertinent features can be compressed, thereby reducing effective dimensionality, and reshaping the manifold to focus on task-relevant features (Ganguli & Sompolinsky, 2012; Rajan et al., 2010; Farrell et al., 2022; Bernardi et al., 2020). Taken together, along the hierarchy of the ventral visual stream, while the representation dimensionality is high and the geometry transformation is homeomorphous, the amount of information in each dimension and the relational structure varies

(Stringer et al., 2019; Lin & Kriegeskorte, 2023). Thus, specialized neuronal modules gradually emerge, within which neurons become increasingly selective and attuned to task demands. This dynamic transformation orchestrates the progression from perceptual to decision-making manifolds, and ultimately to action-oriented manifolds, enabling us to act upon the physical world, in response to stimuli that have acted upon us.

Supplemental Information

Supplemental Materials includes 18 figures, 2 tables and 2 videos.

Acknowledgments: This study was funded by Beijing Municipal Science & Technology Commission, Administrative Commission of Zhongguancun Science Park (Z221100002722012), Tsinghua University Guoqiang Institute (2020GQG1016), and Beijing Academy of Artificial Intelligence (BAAI). We thank Dr. Xingyu Liu for her valuable discussion and comments, and Dr. Haidong Lu for his supervision on conducting the experiments.

Author contributions: H.M., L.J., T.L. and J.L. developed the theoretical framework. H.M. performed the experiments, collected the data, and compiled the data. H.M., L.J., T.L., and J.L. conceived the data analyses. H.M. and L.J. performed the data analyses. L.J. developed the double-twist model. T.L. performed the network simulation. H.M., L.J., and J.L. wrote the paper.

Competing interests: Authors declare no competing interests.

Data and materials availability: The data and analysis code are available at Mendeley Data, https://doi.org/10.17632/vrj3645vgz.1. All data information that supports the findings of this

study are available from the corresponding author on reasonable request.

Reference

Bao, P., She, L., McGill, M., & Tsao, D. Y. (2020). A map of object space in primate inferotemporal cortex. *Nature*, *583*(7814), 103–108. https://doi.org/10.1038/s41586-020-2350-5

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), 954-967.e21. https://doi.org/10.1016/j.cell.2020.09.031

Braddick, O. (1993). Segmentation versus integration in visual motion processing. *Trends in Neurosciences*, *16*(7), 263–268. https://doi.org/10.1016/0166-2236(93)90179-P

B. Schölkopf and A. J. Smola, Learning with Kernels: Support Vector Machines Regularization Optimization and Beyond, Cambridge, MA, USA: *MIT Press*, 2002.

Bullinaria, J. A. (2007). Understanding the emergence of modularity in neural systems. *Cognitive Science*, *31*(4), 673–695. https://doi.org/10.1080/15326900701399939

Caporale, N., & Dan, Y. (2008). *Spike Timing – Dependent Plasticity: A Hebbian Learning Rule*. https://doi.org/10.1146/annurev.neuro.31.060407.125639

Chen, M., Li, P., Zhu, S., Han, C., Xu, H., Fang, Y., Hu, J., Roe, A. W., & Lu, H. D. (2014). An Orientation Map for Motion Boundaries in Macaque V2. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhu235

Chung, S., Lee, D. D., & Sompolinsky, H. (2018). Classification and Geometry of General Perceptual Manifolds. *Physical Review X*, *8*(3), 31003. https://doi.org/10.1103/PhysRevX.8.031003

Chung, S., & Abbott, L. F. (2021). Neural population geometry: An approach for understanding biological and artificial neural networks. *Current Opinion in Neurobiology*, *70*, 137–144. https://doi.org/10.1016/j.conb.2021.10.010

Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Foster, J. D., Nuyujukian, P., Ryu, S. I., Shenoy, K. V. (2012). Neural population dynamics during reaching. *Nature*, *487*(7405), 51–56. https://doi.org/10.1038/nature11129

C. K. Williams and C. E. Rasmussen, Gaussian Processes for Machine Learning, Cambridge, MA, USA: *MIT Press*, 2006.

Cover, T. M. (1965). Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition. *IEEE Transactions on Electronic Computers*, *EC-14*(3), 326–334. https://doi.org/10.1109/PGEC.1965.264137

Cristianini, Nello, and John Shawe-Taylor. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. *Cambridge university press*, 2000.

Dehaqani, M. R. A., Vahabie, A. H., Kiani, R., Ahmadabadi, M. N., Araabi, B. N., & Esteky, H. (2016). Temporal dynamics of visual category representation in the macaque inferior temporal cortex. *Journal of Neurophysiology*, *116*(2), 587–601. https://doi.org/10.1152/jn.00018.2016

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, *73*(3), 415–434. https://doi.org/10.1016/j.neuron.2012.01.010

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

Dyballa, L., Rudzite, A. M., Hoseini, M. S., Thapa, M., Stryker, M. P., Field, G. D., & Zucker, S. W. (2023). *Population encoding of stimulus features along the visual hierarchy*.

Ebitz, R. B., & Hayden, B. Y. (2021). The population doctrine in cognitive neuroscience. *Neuron*, *109*(19), 3055–3068. https://doi.org/10.1016/j.neuron.2021.07.011

Farrell, M., Recanatesi, S., Moore, T., Lajoie, G., & Shea-Brown, E. (2022). Gradient-based learning drives robust representations in recurrent neural networks by balancing compression and expansion. *Nature Machine Intelligence*, *4*(6), 564–573. https://doi.org/10.1038/s42256-022-00498-0

Flesch, T., Juechems, K., Dumbalska, T., Saxe, A., & Summerfield, C. (2022). Orthogonal representations for robust context-dependent task performance in brains and neural networks. *Neuron*, *110*(7), 1258-1270.e11. https://doi.org/10.1016/j.neuron.2022.01.005

Froudarakis, E., Cohen, U., Diamantaki, M., Walker, E. Y., Reimer, J., Berens, P., Sompolinsky, H., & Tolias, A. S. (2020). Object manifold geometry across the mouse cortical visual hierarchy. *BioRxiv*, 2020.08.20.258798. https://doi.org/10.1101/2020.08.20.258798

Fusi, S., Miller, E. K., & Rigotti, M. (2016). Why neurons mix: High dimensionality for higher cognition. *Current Opinion in Neurobiology*, *37*, 66–74. https://doi.org/10.1016/j.conb.2016.01.010

Gallego, J. A., Perich, M. G., Miller, L. E., & Solla, S. A. (2017). Neural Manifolds for the Control of Movement. *Neuron*, *94*(5), 978–984. https://doi.org/10.1016/j.neuron.2017.05.025

Gallego, J. A., Perich, M. G., Naufel, S. N., Ethier, C., Solla, S. A., & Miller, L. E. (2018). Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. *Nature Communications*, *9*(1), 1–13. https://doi.org/10.1038/s41467-018-06560-z

Ganguli, S., & Sompolinsky, H. (2012). Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Annual Review of Neuroscience*, *35*, 485–508. https://doi.org/10.1146/annurev-neuro-062111-150410

Gardner, R. J., Hermansen, E., Pachitariu, M., Burak, Y., Baas, N. A., Dunn, B. A., Moser, M. B., & Moser, E. I. (2022). Toroidal topology of population activity in grid cells. *Nature*, *602*(7895), 123–128. https://doi.org/10.1038/s41586-021-04268-7

Genkin, M., Shenoy, K. V, Chandrasekaran, C., & Engel, T. A. (2023). The dynamics and geometry of choice in premotor cortex. BioRxiv, 2023.07.22.550183 https://www.biorxiv.org/content/10.1101/2023.07.22.550183v1

Gibson, J. J. (1979). *The ecological approach to visual perception. Houghton, Mifflin, and Company.*

Goldstein, E. B., & Cacciamani, L. (2021). *Sensation and perception*. Cengage Learning.

Hu, J., Ma, H., Zhu, S., Hu, J., Ma, H., Zhu, S., Li, P., Xu, H., Fang, Y., Chen, M., & Han, C. (2018). Visual Motion Processing in Macaque V2 Article Visual Motion Processing in Macaque V2. *CellReports*, *25*(1), 157-167.e5. https://doi.org/10.1016/j.celrep.2018.09.014

Jazayeri, M., & Ostojic, S. (2021). Interpreting neural computations by examining intrinsic and embedding dimensionality of neural activity. *Current Opinion in Neurobiology*, *70*, 113–120. https://doi.org/10.1016/j.conb.2021.08.002

Kaufman, M. T., Benna, M. K., Rigotti, M., Stefanini, F., Fusi, S., & Churchland, A. K. (2022). The implications of categorical and category-free mixed selectivity on representational geometries. *Current Opinion in Neurobiology*, *77*(October), 102644. https://doi.org/10.1016/j.conb.2022.102644

Kriegeskorte, N., & Wei, X. X. (2021). Neural tuning and representational geometry. *Nature Reviews Neuroscience*, *22*(11), 703–718. https://doi.org/10.1038/s41583-021-00502-3

Knuth, Donald E. The Art of Computer Programming, Vol. 2: Seminumerical Algorithms, Vol. 2. *Addison-Wesley*, Massachusetts, USA, 1973.

Lamme, V. A. F. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *Journal of Neuroscience*, *15*(2), 1605–1615. https://doi.org/10.1523/jneurosci.15-02-01605.1995

Langdon, C., Genkin, M., & Engel, T. A. (2023). A unifying perspective on neural manifolds and circuits for cognition. *Nature Reviews Neuroscience*, *24*(6), 363–377. https://doi.org/10.1038/s41583-023-00693-x

Lin, B., & Kriegeskorte, N. (2023). The Topology and Geometry of Neural Representations. *Arxiv*. https://doi.org/10.48550/arXiv.2309.11028

Lindsay, G. W., Rigotti, M., Warden, M. R., Miller, E. K., & Fusi, S. (2017). Hebbian learning in a random network captures selectivity properties of the prefrontal cortex. *Journal of Neuroscience*, *37*(45), 11021–11036. https://doi.org/10.1523/JNEUROSCI.1222-17.2017

Low, I. I. C., Giocomo, L. M., & Williams, A. H. (2023). *Remapping in a recurrent neural network model of navigation and context inference*. *BioRxiv*, 2023.01.25.525596 https://doi.org/10.1101/2023.01.25.525596

Lu, H. D., Chen, G., Tanigawa, H., & Roe, A. W. (2010). A Motion Direction Map in Macaque V2. *Neuron*, *68*(5), 1002–1013. https://doi.org/10.1016/j.neuron.2010.11.020

Ma, H., Li, P., Hu, J., Cai, X., Song, Q., & Lu, H. D. (2021). Processing of motion-boundary orientation in Macaque V2. *ELife*, *10*, 1–24. https://doi.org/10.7554/eLife.61317

Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, *503*(7474), 78–84. https://doi.org/10.1038/nature12742

Marcar, V. L., Raiguel, S. E., Xiao, D., & Orban, G. a. (2000). Processing of kinetically defined boundaries in areas V1 and V2 of the macaque monkey. *Journal of Neurophysiology*, *84*(6), 2786–2798.

Minsky, M. L., & Papert, S. (1969). Perceptrons, Expanded Edition, An Introduction to Computational Geometry. In *MIT Press*.

Muller, M. E. (1959). A note on a method for generating points uniformly on n-dimensional spheres. *Communications of the ACM*, *2*(4), 19–20. https://doi.org/10.1145/377939.377946

Mysore, S. G., Vogels, R., Raiguel, S. E., & Orban, G. a. (2006). Processing of kinetic boundaries in macaque V4. *Journal of Neurophysiology*, *95*(3), 1864–1880. https://doi.org/10.1152/jn.00627.2005

Nogueira, R., Rodgers, C. C., Bruno, R. M., & Fusi, S. (2023). The geometry of cortical representations of touch in rodents. *Nature Neuroscience*, *26*(2), 239–250. https://doi.org/10.1038/s41593-022-01237-9

Okazawa, G., Hatch, C. E., Mancoo, A., Machens, C. K., & Kiani, R. (2021). Representational geometry of perceptual decisions in the monkey parietal cortex. *Cell*, *184*(14), 3748-3761.e18. https://doi.org/10.1016/j.cell.2021.05.022

Perich, M. G., Conti, S., Badi, M., Bogaard, A., Barra, B., Wurth, S., Bloch, J., Courtine, G., Micera, S., Capogrosso, M., & Milekovic, T. (2020). Motor cortical dynamics are shaped by multiple distinct subspaces during naturalistic behavior. *BioRxiv*, 2020.07.30.228767. https://www.biorxiv.org/content/10.1101/2020.07.30.228767v2

Poland, J. (2000). Three different algorithms for generating uniformly distributed random points on the N-sphere. *Computer Science, Physics.* http://www-alg.ist.hokudai.ac.jp/~jan/randsphere.pdf.

Rajan, K., Abbott, L. F., & Sompolinsky, H. (2010). Stimulus-dependent suppression of chaos in recurrent neural networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *82*(1). https://doi.org/10.1103/PhysRevE.82.011903

Rigotti, M., Barak, O., Warden, M. R., Wang, X. J., Daw, N. D., Miller, E. K., & Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature*, *497*(7451), 585–590. https://doi.org/10.1038/nature12160

Russo, A. A., Khajeh, R., Bittner, S. R., Perkins, S. M., Cunningham, J. P., Abbott, L. F., & Churchland, M. M. (2020). Neural Trajectories in the Supplementary Motor Area and Motor Cortex Exhibit Distinct Geometries, Compatible with Different Classes of Computation. *Neuron*, *107*(4), 745-758.e6. https://doi.org/10.1016/j.neuron.2020.05.020

Seung HS & Lee DD. (2000). The manifold ways of perception. *Science 290: 2268–2269.*

She, A. L., Benna, M. K., Shi, Y., Fusi, S., & Tsao, D. Y. (2021). The geometry of face memory. *bioRxiv*,2021.03.12.435023. https://doi.org/10.1101/2021.03.12.435023

Sporns, O. (2014). Contributions and challenges for network models in cognitive neuroscience. *Nature Neuroscience*, *17*(5), 652–660. https://doi.org/10.1038/nn.3690

Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M., & Harris, K. D. (2019). High-dimensional geometry of population responses in visual cortex. *Nature*, *571*(7765), 361–365. https://doi.org/10.1038/s41586-019-1346-5

Weigand, M., Sartori, F., & Cuntz, H. (2017). Universal transition from unstructured to structured neural maps. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(20), E4057–E4064. https://doi.org/10.1073/pnas.1616163114

Xie, Y., Hu, P., Li, J., Chen, J., Song, W., Wang, X. J., Yang, T., Dehaene, S., Tang, S., Min, B., & Wang, L. (2022). Geometry of sequence working memory in macaque prefrontal cortex. *Science*, *375*(6581), 632–639. https://doi.org/10.1126/science.abm0204

Materials and Methods

Four hemispheres from two adult male macaque monkeys (Macaca mulatta) were used in this study. All procedures were performed in accordance with the National Institutes of Health Guidelines and were approved by the Institutional Animal Care and Use Committee of the Beijing Normal University (protocol number: IACUC(BNU)-NKCNL2013-13).

Visual stimuli, behavioral tasks, and recordings

The data used in this study was from our previous work, and for details on stimuli, tasks, and recordings, please see Ma et al. (2021).

Stimuli were generated with ViSaGe using MATLAB scripts and presented on a 21-inch CRT display. MIC stimuli had seven levels of dot motion coherence. The size of the MIC was a 4˚-diameter circular and the position of MIC was consistent with array population RF.

Two monkeys performed an MIC orientation-discrimination task after headpost and optical imaging guided 32-channel Utah array implant. This task was a two-alternative forced-choice discrimination task and the monkeys were trained to make an eye saccade choice based on MIC orientations. The monkeys made a saccade to the right target if the orientation was tilted to the right of the vertical axis, and vice versa. The monkeys received water reward for correct choices.

The electrophysiological recording system is AlphaLab SnR 64-channel system. Neural signals were sampled at 22 kHz and with an 800 – 7500 Hz bandpass filter. In this study, we only used single neurons in unique-unit dataset (Ma et al., 2021).

Data analysis

Data preprocessing. To build a high-dimensional neural space, we identified all V2 neurons that participated in both the MIC orientation-discrimination task and the MIC fixation task. In total, we obtained 93 V2 single neurons, with 47 neurons from Monkey S and 46 neurons from Monkey W.

We sorted all trials into 112 conditions to analyze each neuron's response (2 motion-axis orientation conditions, 4 motion-shear conditions, 7 coherence levels, and 2 performance outcomes). Subsequently, we

calculated each neuron's trial-averaged response (from -200ms before stimulus onset to 200ms after stimulus offset) with a Gaussian window (a 10ms sliding window with a 2ms step size). We then combined all neuron responses after z-scoring each neuron's trial average response (Mante et al., 2013; Okazawa et al., 2021). Additionally, we excluded conditions with fewer than 3 trials for some neurons. Following these steps, we constructed a data matrix of dimensions 61 (useful conditions) x T (trial time) x 93 (neuron number) to represent the data from the MIC orientation discrimination task.

Support vector machine (SVM). We used SVM for two purposes: first, to decode categorical information from the neural data, and second, to provide a well-defined vector known as the direction of interest (DoI), which represents a distinct dimension in the neural space. In the analysis, we used the *fitcsvm* function with a linear kernel in MATLAB. We retained the default hyper-parameter values of the function, except for customizing the box constraint. In the neural geometry part, the box constraint was set at 0.001, whereas for the binary classification part, it was set at 1 (for an explanation of box constraints, see Supplementary Methods M.1). Then, the SVMs were trained and tested using data points in the neural space.

We divided the data points into training and testing sets. First, we partitioned the complete time range from -200ms to 1000ms relative to stimulus onset into tiled 12ms-wide time bins. Within each time bin, we randomly selected half of the time steps for training and the other half for testing. To train a single SVM classifier, the selected training data from time bins were pooled together. To generate time courses of classifier accuracy, the classifier was tested within each time bin using the testing data.

Angle analysis. To calculate the angle subtended between two n-dimensional unit vectors $\boldsymbol{\beta}_i$ and $\boldsymbol{\beta}_j$, we used the following formula:

$$\theta = \mathrm{acos}\left(\boldsymbol{\beta}_i^{\mathrm{T}} \boldsymbol{\beta}_j\right).$$

The vectors were the DoI from SVM classification. For the neural geometry and twist model parts, we performed SVM analysis once for each classification. For control analyses, we ran SVM analysis 200 times for each classification (see Supplementary Materials S.3 and Fig. S3A&B). The orthogonality of the angle was tested by examining whether it significantly differed from the angle distribution constructed by randomly

selecting two vectors in a 93-dimensional space (see Supplementary Materials S.3 and Fig. S3C).

Three-D visualization. The main axes identified by SVM were mutually orthogonal, resulting in Cartesian coordinate systems. Then, we linearly transformed the original neural space to a new coordinate system where its first several axes aligned with the main axes identified by SVM. To obtain the transformation matrix $\boldsymbol{T}$, we first created an arbitrary full rank matrix $\boldsymbol{A}$, where its first few columns were the identified main axes. We then applied QR decomposition on matrix $\boldsymbol{A}$ to obtain an orthogonal matrix $\boldsymbol{Q}$. The transformation matrix was $\boldsymbol{T} = \boldsymbol{Q}^{\mathrm{T}}$ (see Supplementary Methods M.2 for details).

To visualize the neural states in a 3-D subspace, we first applied low-pass filter (cutoff frequency: 2Hz) to the data in the original neural space. We then used matrix $\boldsymbol{T}$ to transform the filtered data into a new high-dimensional coordinate system. Finally, the transformed data were projected into a 3-D subspace constructed by the main axes identified for visualization.

Double-twist model. The double-twist model transformed the continuous cylindrical stimulus manifold depicted in Fig. 3C (left) to a continuous perceptual manifold embedded in a 7-D space. Its projections into 3-D subspaces were shown in gray in Fig. 3C (middle & right). The perceptual manifold arose from the correspondence between the XOR operation and the arithmetic product. Let the true value be represented by -1.0, and the false value by 1.0. The truth table of the XOR operation aligns with that of the arithmetic product (see Fig. M1). That is,

$$x \oplus y \Leftrightarrow xy.$$

Therefore, in the double-twist model, the 7-D coordinates of a point in the perceptual manifold corresponding to a point $[x, y, z]$ in the stimulus manifold were $[x, y, z, xy, yz, xz, xyz]$.

The double-twisting model was fit to the neural data using affine transformation. Because the neural data contained noise, we first applied a low-pass filter (cutoff frequency: 2Hz), and then calculated the steady-state averages (from 300ms to 500ms relative to stimulus onset) of neural activities for the 8 stimuli to determine 8 centers. These centers were then projected into various 3-D subspaces (Fig. 3). In each subspace, we located the neural states for the 8 stimuli based on the ideal double-twisting model. Using the least-square

method, we derived the transformation matrix $\boldsymbol{F}$. The mean of the residuals was represented as a vector $\boldsymbol{e}$. For any point $\boldsymbol{x}$ on the ideal double-twist model, we applied the transformation

$$\hat{\boldsymbol{y}} = \boldsymbol{F}\boldsymbol{x} + \boldsymbol{e}$$

to fit the model to the data (see Supplementary Methods M.5 for detailed formulation). The goodness of fit was measured by R^2.

Binary classification. Binary classification entailed sorting stimulus conditions into two classes based on all possible classification rules. With 8 stimulus conditions, we had a total of 256 classification rules. For each classification, the criterion for linear separability was set at 75% accuracy within each stimulus cluster. Each neuron pool was randomly selected from the entire set of 93 neurons, one at a time, and we then conducted SVM analysis 10 times for each neuron pool (see Supplementary Methods M.6).

To calculate the dimensionality of the network, we employed the approach developed by Fusi et al. (2013). Briefly, we calculated the ratio of the actual number of linearly separable classifications and the theoretical number of binary classifications for each condition number n ($n \in \{2, \dots, 8\}$). The last n whose ratio > 0.8 was selected as the dimensionality of the network (see Supplementary Methods M.6).

Connectivity patterns of neural networks. We created a simple two-layer feedforward neural network; the first layer contained 3 units representing the three feature dimensions, respectively, and the second layer contained 93 nonlinear mixed selectivity neurons. Each neuron received the input signals of all the three feature dimensions (mixed): x(HV), y(OI) and z(CA), and employed a ReLU activation function (nonlinearity). The input signals represented the eight corners of the cube in the x-y-z stimulus space, each corresponding to a specific stimulus. The activity r_i of the ith neuron was defined as:

$$r_i = \mathrm{ReLU}\left(\sum_j w_{ij}\gamma_j + b_i\right) + \epsilon,$$

where $\gamma_j \in \{x, y, z\}$ is the input signal of the jth stimulus feature with w_{ij} as the weight, b_i is a random bias sampled from the uniform distribution $\mathcal{U}(0,1)$, and ϵ is a noise term drawn from the Gaussian distribution $\mathcal{N}(0, 1/3)$.

The weights w_{ij} were randomly sampled from standard Gaussian distributions. The covariance between the weights controlled the structure of the network and thus the heterogeneity of the output activities. Hence, we used a diversity parameter $d \in [0,1]$ to define the 93×93-D covariance matrix K for sampling the weights of feature x:

$$K = \begin{bmatrix} 1 & 1-d & \cdots & 1-d \\ 1-d & 1 & \cdots & 1-d \\ \vdots & \vdots & \ddots & 1-d \\ 1-d & 1-d & \cdots & 1 \end{bmatrix}.$$

The covariance matrix K was fed into the *multivariate_normal* function in Python's *numpy.random* package to generate the weights. The procedure for generating weights of features y and z was identical.

We ran multiple simulations on the network. In each simulation, input stimuli were repeated 100 times to allow the added random noise for generating point clusters. We applied the same analyses for the network's output activities as we did for the neural data. One hundred simulations were conducted for calculating the dimensionality of the neural geometry.