



# A neural geometry approach comprehensively explains apparently conflicting models of visual perceptual learning

Received: 19 May 2024

Accepted: 20 February 2025

Published online: 31 March 2025

Check for updates

Yu-Ang Cheng <sup>1,2</sup>, Mehdi Sanayei <sup>3,4</sup>, Xing Chen <sup>5</sup>, Ke Jia <sup>6,7,8</sup>, Sheng Li <sup>9,10,11</sup>, Fang Fang <sup>9,10,11,12</sup>, Takeo Watanabe <sup>12</sup>, Alexander Thiele <sup>3</sup> & Ru-Yuan Zhang <sup>1</sup>✉

Visual perceptual learning (VPL), defined as long-term improvement in a visual task, is considered a crucial tool for elucidating underlying visual and brain plasticity. Previous studies have proposed several neural models of VPL, including changes in neural tuning or in noise correlations. Here, to adjudicate different models, we propose that all neural changes at single units can be conceptualized as geometric transformations of population response manifolds in a high-dimensional neural space. Following this neural geometry approach, we identified neural manifold shrinkage due to reduced trial-by-trial population response variability, rather than tuning or correlation changes, as the primary mechanism of VPL. Furthermore, manifold shrinkage successfully explains VPL effects across artificial neural responses in deep neural networks, multivariate blood-oxygenation-level-dependent signals in humans and multiunit activities in monkeys. These converging results suggest that our neural geometry approach comprehensively explains a wide range of empirical results and reconciles previously conflicting models of VPL.

Adapting to new visual environments is crucial for an organism's survival in its environment. This ability is well exemplified by visual perceptual learning (VPL), which is defined as long-term performance enhancements resulting from visual experience<sup>1,2</sup>. However, despite years of research in systems neuroscience, psychophysics and machine learning, the mechanisms behind VPL remain mysterious.

It is widely acknowledged that visual training enhances behavioural performance and refines representations in neural populations. Previous studies using human neuroimaging and monkey neurophysiology have demonstrated a significant improvement in the fidelity of stimulus encoding within population responses<sup>3–5</sup>. These findings strongly support the theory that enhanced signal-to-noise ratios (SNRs) serve as a potent computational mechanism for improved neural

representations associated with VPL (Fig. 1e)<sup>6–8</sup>. However, improved SNR is an algorithm-level model, and the exact underlying neural mechanisms to achieve improved SNR remain elusive. Several conflicting models have been proposed on the basis of neural changes associated with VPL. One model suggests that VPL is associated with changes in population representations resulting from changes in neuronal tuning curves, as indicated by sharpened orientation tuning curves in monkey visual cortex<sup>9,10</sup>. Another model assumes that changes in population representations result from a reduction in trial-by-trial co-variation of neuronal firing rate, known as noise correlations, which have been observed in association with VPL in both monkeys and songbirds<sup>11–14</sup>.

The primary conceptual gap in reconciling the conflicting models of VPL lies in their focus on mechanisms proposed at the single-unit

level (for example, changes in tuning curves or noise correlations), whereas the effects of these mechanisms in VPL should be evaluated at the population level (that is, improved population representations). Although VPL is indeed associated with changes in both single-unit responses and improved population representations, it remains controversial whether changes in single-unit responses are the actual cause or merely by-products of improved population representations. While these conflicting models of VPL capture certain aspects of the empirical findings, they fail to generate falsifiable predictions about how changes in single-unit responses contribute to improved population representations.

A major obstacle to comparing the conflicting models of VPL is the complex interactions between different aspects of single-unit responses (for example, tuning curves and noise correlations) on population representations. Computational neuroscience research has elucidated that the impact of noise correlations on population representations heavily depends on its interaction with tuning curves<sup>15</sup>. It is important to note that reduced noise correlations do not inherently enhance information in a neural population<sup>16–18</sup>. Moreover, the challenge is exacerbated by the fact that their interaction effects are even changing rather than remaining stable throughout a training process. These dynamic changes further complicate the understanding of how training affects their interactions. To overcome this, a comprehensive computational approach is imperative to quantify and disentangle the effects of different changes in single-unit responses, such as sharpened tuning curves and reduced noise correlations, on neural representations at the population level.

To comprehensively explain these conflicting models, we developed a neural geometry approach of VPL. In this approach, trial-by-trial population responses elicited by two stimuli for discrimination form two differentiable manifolds in a high-dimensional neural space. In this space, changes in single-unit responses (for example, tuning curves, Fano factor and noise correlations) can be interpreted as changes in several fundamental and measurable geometric properties (for example, centroids, size and orientations) of neural manifolds. This approach allows quantitative comparisons of conflicting models of VPL and assessments of their contributions to population representations within the same computational framework. Thus, this approach directly bridges single-unit responses and population representations and offers a normative account of the potential neural mechanisms underlying VPL. Specifically, this approach proposes four possible training-induced geometric changes (signal enhancement, manifold shrinkage, signal rotation and manifold warping) that can summarize all previous models of VPL. Thus, improved population representations can be achieved by one or a combination of the four interpretable mechanisms.

Our study includes theoretical modelling and empirical tests of model predictions. First, to assess this neural geometry approach, we trained deep convolutional neural networks (DCNNs) on the typical VPL task—orientation discrimination learning—and found that the DCNNs successfully replicated a wide range of psychophysical and imaging findings in humans, as well as neurophysiological findings in monkeys. Second, analyses of the geometric mechanisms mentioned above suggest that changes in both tuning curves and noise correlations are indeed present in VPL. Third, and most importantly, our analysis further revealed that neither changes in tuning curves nor changes in noise correlations at the single-unit level contributed significantly to improved population representations. Surprisingly, we found that neural manifold shrinkage induced by reduced response variability emerged as the primary mechanism driving VPL. Our neural geometry approach generates several empirical testable predictions. We directly tested these predictions on empirical data across different tasks, different levels of measurement and different brain regions in different species. Remarkably, we found that the geometry approach incorporating manifold shrinkage aligned closely with the activity of artificial neurons in DCNNs trained on VPL of motion direction discrimination

learning task, blood-oxygenation-level-dependent (BOLD) response changes associated with VPL of motion direction learning in humans, and the electrophysiological population response changes associated with VPL of contrast discrimination in monkey V4.

## Results

### VPL improves behavioural performance of DCNN

To elucidate the neurocomputational mechanisms of VPL, we trained a DCNN (Fig. 1a) to perform a classical orientation discrimination task<sup>7</sup>. DCNN modelling allows us to easily assess the activity of the whole population in each layer and along the entire visual hierarchy. Similar to the neural network in ref. 19, this neural network inherits the first five convolutional layers of AlexNet, which was pretrained on ImageNet<sup>20</sup>. To emulate the decision stage of orientation discrimination, we added a linear decoding layer and used the logistic function to classify the activity of the decision unit into a binary perceptual choice (that is, clockwise or counterclockwise rotation of the target stimulus relative to the reference stimulus). Importantly, similar to previous psychophysical studies<sup>7,21</sup>, we systematically manipulated the level of input image noise (Fig. 1b). The network was trained on stimuli with multiple noise and contrast levels (see Methods for training details).

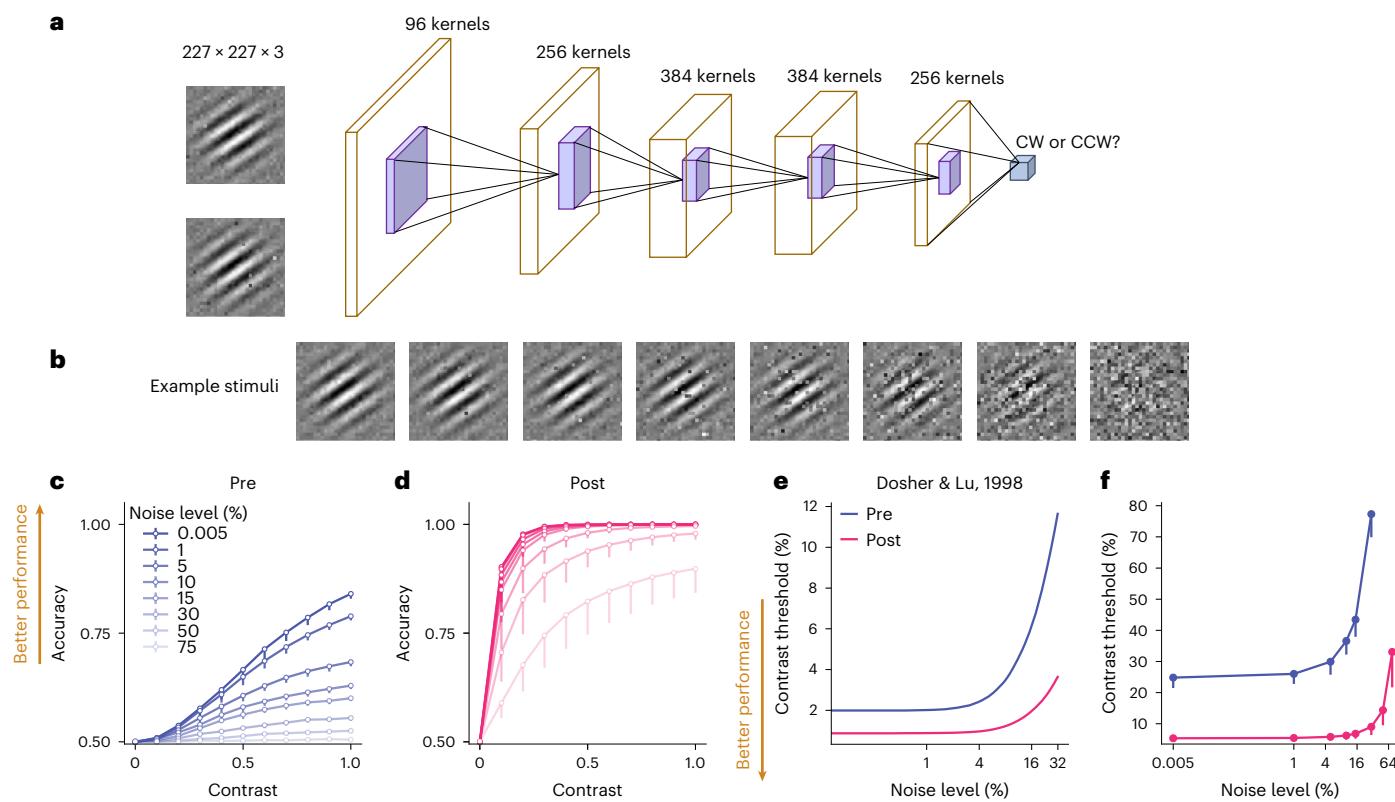
To evaluate the performance of the neural network, we assessed orientation discrimination accuracy as a function of stimulus contrast and noise (Fig. 1c,d) and further derived contrast thresholds as a function of image noise level (Fig. 1f, threshold versus noise (TvN) function). We found that training improved the network performance in this task in almost all stimulus contrast and noise conditions. The uniform downshift of TvN functions (Fig. 1f) is consistent with well-established human psychophysical results (replotted in Fig. 1e)<sup>7,8</sup>.

### VPL refines neural population representations in DCNN

We next sought to understand the effects of visual training on population representations in the network. We performed multivariate decoding analyses in each layer and found that training significantly improved decoding accuracy in later layers (Fig. 2f, layers 3–5; one-sided paired *t*-test, all  $t_{(3)} < -3.59$ , all  $P < 0.020$ ; see full statistical results in Supplementary Table 1). More formally, we calculated linear Fisher information, a classical metric in computational neuroscience, to quantify how well the two stimuli can be discriminated on the basis of population responses (Methods). The amount of sensory information represented in later layers was indeed significantly enhanced by training (Fig. 2g, layers 3–5; one-sided paired *t*-test, all  $t_{(3)} < -3.47$ , all  $P < 0.018$ ; see full statistical results in Supplementary Table 2). Such refined neural representation at the population level is consistent with the decoding results based on both cortical activity in humans<sup>3,4,22</sup> and multiunit spiking activity in monkeys<sup>13,14</sup>.

### VPL changes response properties of individual units in DCNN

In addition to the population-level changes, we found that three key individual-level neural signatures of VPL as documented in the neurophysiological literature emerge naturally from the neural network training. First, training modestly sharpened the tuning curves of artificial neurons in layers 1–4 (Fig. 2h and Supplementary Fig. 1), a finding reported in several previous studies<sup>9,10,23</sup> (Fig. 2c, but see also null results in ref. 24). Second, we observed a decrease in Fano factor of individual units in all five layers (Fig. 2i and Supplementary Fig. 1), a phenomenon indicating an increased SNR of individual neuronal responses in both humans<sup>25</sup> and monkeys<sup>11,23</sup> (Fig. 2d). The sharpened tuning curve and reduced Fano factor are also consistent with theoretical modeling<sup>19</sup>. Third, training reduced trial-by-trial noise correlations between units in all five layers (Fig. 2j and Supplementary Fig. 1), a finding also consistent with several empirical results in monkeys<sup>11–14</sup>. Critically, we also found that the reduction in noise correlation depended on tuning similarity. Learning reduced the noise correlations between units with similar tunings (that is, positive signal correlations) and increased the



**Fig. 1 | DCNN Modelling of orientation VPL.** **a,b**, A DCNN (**a**) is trained on an orientation discrimination task (clockwise, CW or counter-clockwise, CCW) with Gabor stimuli embedded in different levels of image noise (**b**). **c,d**, Orientation discrimination accuracy is improved from pre-test (**c**) to post-test (**d**). **e,f**, Training induces a downshift of the threshold versus noise function (**f**), an effect that is qualitatively similar to existing human psychophysical results

(**e**, corresponds to the 70.7% accuracy condition in fig. 1 of ref. 7). The absolute quantitative differences between **e** and **f** may be due to differences in the overall SNR or the number of layers and units between the human visual system and the DCNN. Data are presented as mean  $\pm$  s.e.m., with error bars and error shadings in **c** and **d** representing the s.e.m. across four ( $n = 4$ ) reference orientations.

noise correlations between units with opposite tunings (that is, negative signal correlations) (Supplementary Fig. 2). Previous theoretical work has suggested that the former type of noise correlations is detrimental for information coding and the latter type is beneficial<sup>15,16</sup>. The pattern of reduced detrimental and increased beneficial noise correlations has been discovered with learning tasks in songbirds<sup>26</sup> and with attention tasks in monkeys<sup>27</sup>.

In addition to these classical neurophysiological findings in VPL, our network also captures some important response properties of sensory neurons in the primate early visual system. First, the relationship between the Fano factor and orientation tuning of the artificial neurons bears strong resemblances to the empirical measures of V1 neurons in monkeys<sup>28</sup> (Supplementary Fig. 2). Second, we found a positive relationship between signal correlation and noise correlation among artificial neurons in all layers (Supplementary Fig. 2). This relationship has also recently been documented as a ubiquitous phenomenon in both electrophysiological<sup>29–31</sup> and human imaging<sup>17,18,32</sup> studies.

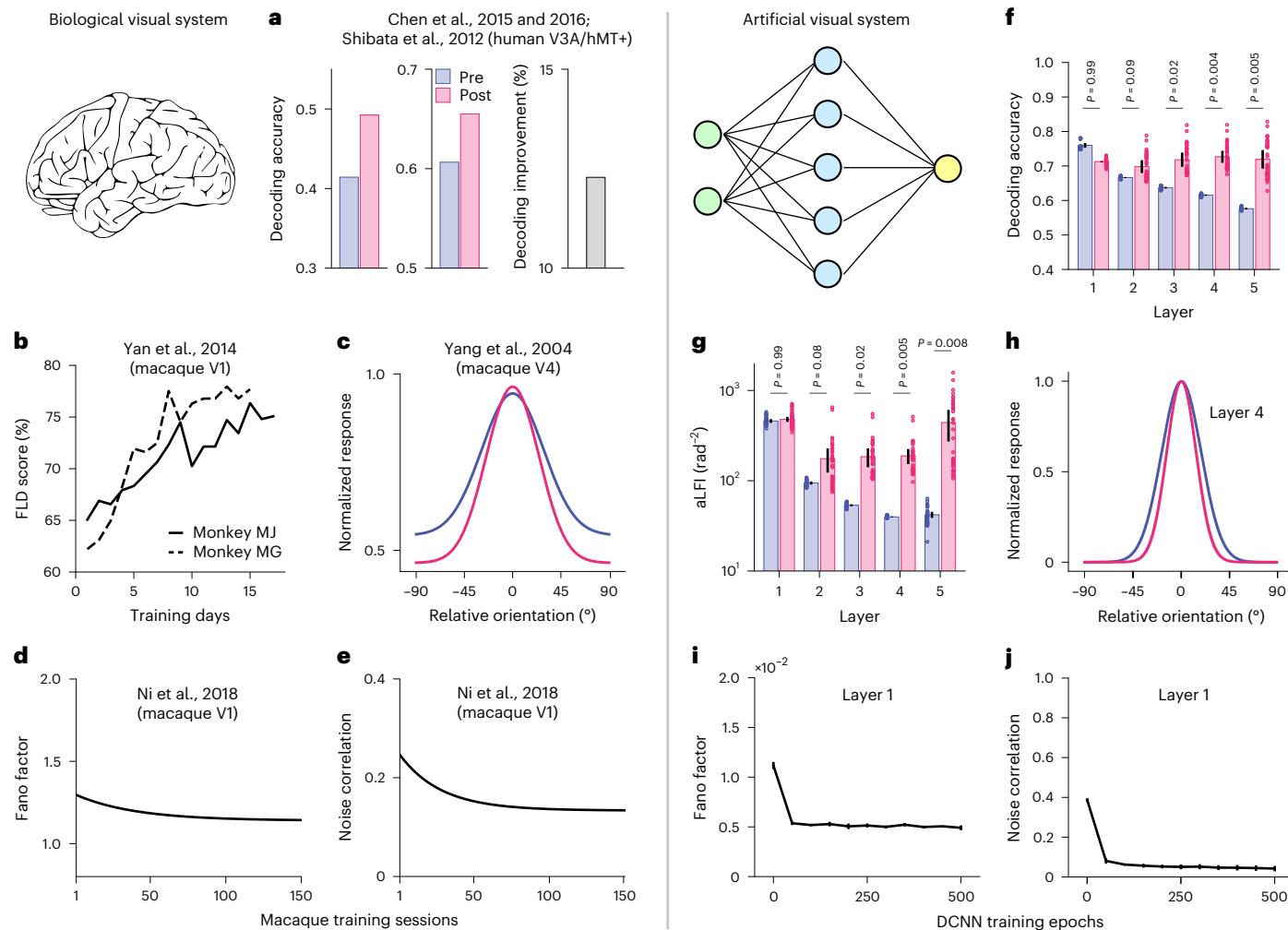
Taken together, these results suggest that our DCNNs are powerful models and allow us to explore neurocomputational mechanisms that may be difficult to elucidate in empirical experiments. Here we focus on the qualitative similarities of learning-induced changes in DCNN and in certain brain regions. However, we did not attempt to claim one-on-one mapping between DCNN layers and brain regions because this requires one to build precise encoding models.

#### Four mechanisms and the neural geometry approach of VPL

How would improved sensory discrimination manifest in high-dimensional population responses? In the simplified one-dimensional scenario (Fig. 3a), the classical signal detection theory posits that

better sensory discrimination can be achieved by either increasing the distance between the means (that is, signal enhancement) and/or decreasing the variance (that is, noise reduction) of the two response distributions. In multivariate population responses, the two stimuli to be discriminated instead generate two multivariate response distributions (that is, neural manifold) in a high-dimensional neural space whose dimension corresponds to the number of units in a population (Fig. 3b,c). In a simplified visualization in a two-dimensional space (Fig. 3d), the two distributions are elliptical due to noise correlations between units. We refer to the vector connecting the mean of the two distributions as the signal vector and its modulus length (that is, the Euclidean distance between the two manifold centroids) as the signal separation.

In the high-dimensional neural space, our neural geometry approach of VPL proposes that visual training improves sensory discrimination by shaping some fundamental geometric properties of the neural manifolds. Here, under this approach, there exist only four possible mechanisms to further separate two neural manifolds (equation (4) in Methods). First, according to the classical signal detection theory, the signal enhancement mechanism predicts an increased Euclidean distance between the centroids of the two neural manifolds (Fig. 3e). However, we found that the signal separation between the two manifolds did not significantly increase with learning in all five layers, and even slightly decreased in the first two layers (Fig. 3f; one-sided paired  $t$ -test, all  $t_{(3)} > -1.27$ , all  $P > 0.146$ , all Bayes factor  $BF_{10} < 1.46$ ; see full statistical results in Supplementary Table 3). Second, the manifold shrinkage mechanism predicts that visual training reduces the trial-by-trial response variance of units, thereby reducing the size of the manifolds (Fig. 3g). This is what we found in all five layers (Fig. 3h;



**Fig. 2 | DCNN models reproduce empirical findings.** **a–j**, Neural correlates of VPL in humans (**a**), monkeys (**b–e**) and our DCNN (**f–j**). Visual training improves stimulus decoding accuracy in related regions in the human brain (**a**) and decoding scores of Fisher's linear discriminant (FLD) in monkey V1 (**b**). Visual training sharpens orientation tuning curves of neurons in monkey V4 (**c**) and also reduces Fano factors and interneuron noise correlations (**d** and **e**). Similar results are observed in the DCNN: network training also improves decoding accuracy in layers 3–5 (layers 1 and 2: one-sided paired *t*-test,  $t_{(3)} > -1.79$ , all  $P > 0.08$ ; layers 3–5: one-sided paired *t*-test,  $t_{(3)} < -3.59$ , all  $P < 0.02$ ; see full statistical results in Supplementary Table 1; **f**), and aLFI (total information in each layer divided by the number of units in that layer) in layers 3–5 (layers 1 and 2: one-sided paired

*t*-test,  $t_{(3)} > -1.84$ , all  $P > 0.08$ ; layers 3–5: one-sided paired *t*-test, all  $t_{(3)} < -3.47$ , all  $P < 0.02$ ; see full statistical results in Supplementary Table 2; **g**). Training sharpens orientation tuning curves of units in layers 1–4 in the DCNN (results of layer 4 only are shown in **h**). Similar reduction of Fano factors and noise correlations are observed in the DCNN (results of layer 1 only are shown in **i** and **j**). The data shown in **h–j** are the median value across units in a layer. The results of all five layers are shown in Supplementary Fig. 1. Panels **a–e** are reproduced by the data points shown in the original papers. Data are presented as mean  $\pm$  s.e.m., with error bars and error shadings in **f–j** represent the s.e.m. across four ( $n = 4$ ) reference orientations (error shadings in **h** are small and barely visible).

one-sided paired *t*-test, all  $t_{(3)} > 8.39$ , all  $P < 0.002$ ; see full statistical results in Supplementary Table 4). We further included two previously overlooked mechanisms that can only occur in high-dimensional neural space and increase manifold discriminability. In the third mechanism, although visual training did not increase signal separation, it may change the relative positions of the centroids of the two manifolds and consequently increase discriminability due to the elliptical shape of the manifolds (Fig. 3*i*). Interestingly, we found that the signal vectors in each layer were rotated by  $-50\text{--}70^\circ$  after training (Fig. 3*j*). We call this mechanism signal rotation. Fourth, visual training can warp the shapes of the high-dimensional neural manifolds while keeping the size of the manifolds unchanged. As indicated by the change of covariance structure, we found that visual training systematically warped the shape (that is, covariance structures) of the high-dimensional neural manifolds (Fig. 3*k–m*). We refer to this mechanism as manifold warping. Note that manifold warping includes both the changes in correlation

structures and the redistribution of variances across individual units, while holding the total variance constant. It is manifold shrinkage that attenuates the total variance.

**Information-theoretic analyses quantified mechanisms of VPL**  
Given the four possible mechanisms (that is, signal enhancement, manifold shrinkage, signal rotation and manifold warping) and their complex interaction effects, how can we delineate their respective contributions to improved population representations? Here we use linear Fisher information to quantify manifold separability. Besides, we introduce a stepwise approach to further disentangle the respective contributions of the four possible mechanisms. Specially, their respective contributions are assessed by sequentially allowing only one mechanism to occur and quantifying its endowed changes in the linear Fisher information of whole populations (Fig. 4*a*). For example, as shown in Fig. 4, we first calculate how much information is enhanced

by considering only the signal enhancement scenario, then by considering both signal enhancement and manifold shrinkage, and so on until all four mechanisms are included.

Interestingly, we found that the effect of signal enhancement is minimal in all five layers. This mechanism even reduces stimulus information in layers 1 and 2. This is consistent with the reduced Euclidean distance in the first two layers (Fig. 3f). Manifold shrinkage enhances stimulus information in almost all layers. Interestingly, we found that signal rotation appears to enhance stimulus information (Fig. 4b, green bars). This is because rotation of the signal vectors disrupts their relative parallelism to the covariance direction at pre-test, making them more orthogonal. Such changes increase the apparent information. However, the effect becomes minimal when manifold warping is further considered (Fig. 4b, magenta bars) because visual training also warps the covariance direction to realign it with the post-test signal vector, thereby reducing stimulus information (see more explanations in Supplementary Note 1 and full statistical results in Supplementary Table 5).

Taken together, we propose an interpretable and quantitative neural geometry approach of VPL where visual training refines the geometry of representations in a high-dimensional neural space. Using this approach, we found that three of four possible mechanisms occur in VPL. Most importantly, we found that manifold shrinkage in population responses was the key mechanism underlying the improved population representations induced by visual training in the DCNN. However, all above findings are the results of theoretical modelling using our DCNN model of orientation discrimination learning. Several predictions here have never been tested or reported in empirical studies. In the rest of ‘Results’, we tested these predictions across diverse tasks, measurement modalities and species.

### Motion direction discrimination learning in DCNN

The above analyses focus only on one classical VPL task—orientation discrimination and a specific neural network structure—a six-layer convolutional neural network. In this section, we switch to motion VPL—another sensory domain that is also widely used in psychophysical<sup>33,34</sup>, human imaging<sup>3,4</sup> and neurophysiological studies<sup>35</sup>. Importantly, motion VPL involves the processing of both spatial and temporal signals rather than merely static spatial information in orientation learning. Similarly, we inherited the first six layers of the pretrained C3D network<sup>36</sup> and trained the neural network to perform a motion direction discrimination task commonly used in psychophysics (see Methods for stimulus and training details).

In the motion DCNN, we found similar mechanisms as in the orientation discrimination learning task. First, motion direction discrimination training improved the behavioural performance of the network (Fig. 5b). Second, training also enhanced decoding accuracy and averaged linear Fisher information (aLFI) in later layers (Fig. 5c,d;

layers 4–6; for decoding accuracy: one-sided paired *t*-test, all  $t_{(3)} < -7.14$ , all  $P < 0.02$ ; for aLFI: one-sided paired *t*-test, all  $t_{(3)} < -7.22$ , all  $P < 0.003$ ; see full statistical results in Supplementary Tables 6 and 7), suggesting that such training refines stimulus representation at the population level. Third, the effects of motion direction discrimination training on individual units in layer 6 are also pronounced (see results for all six layers in Supplementary Fig. 3). We found that training reduced Fano factor (Fig. 5e; one-sided paired *t*-test,  $t_{(3)} = 57.58$ ,  $P < 0.001$ , one-sided 95% confidence interval (CI)  $3.1 \times 10^{-2}$  to  $\infty$ , Cohen’s  $d = 38.75$ ) and noise correlations (Fig. 5f; one-sided paired *t*-test,  $t_{(3)} = 42.84$ ,  $P < 0.001$ , one-sided 95% CI  $4.4 \times 10^{-4}$  to  $\infty$ , Cohen’s  $d = 2.19$ ). Fourth, training did not significantly improve signal separation (Fig. 5g; one-sided paired *t*-test,  $t_{(3)} = -0.98$ ,  $P = 0.198$ , one-sided 95% CI  $-\infty$  to  $1.7 \times 10^{-2}$ , Cohen’s  $d = -0.19$ ,  $BF_{10} 1.22$ ) but markedly reduced response variance (Fig. 5h; one-sided paired *t*-test,  $t_{(3)} = 59.05$ ,  $P < 0.001$ , one-sided 95% CI  $2.2 \times 10^{-2}$  to  $\infty$ , Cohen’s  $d = 43.89$ ). In addition, motion direction discrimination training also induced two previously overlooked mechanisms: signal rotation (Fig. 5i) and manifold warping (Fig. 5j,k). Most importantly, the four mechanisms induced by the training had similar respective contributions to population representations (Fig. 5l).

### Motion direction discrimination learning in the human brain

The converging results in the DCNNs of orientation and motion direction discrimination, and the remarkable agreement between our DCNNs and existing empirical neuroscientific findings, support the biological plausibility of our DCNNs. However, it remains unknown whether these predictions are present only in the DCNNs and have no biological basis in the brain. To address this question, we analysed BOLD responses in the cortex of human subjects before and after they were trained on a motion direction discrimination task (Fig. 6a, ref. 37). Twenty-two human subjects participated in the motion VPL study. Subjects were trained for 10 days on a fine-direction discrimination task, and psychophysical and functional magnetic resonance imaging (fMRI) tests were performed before and after training.

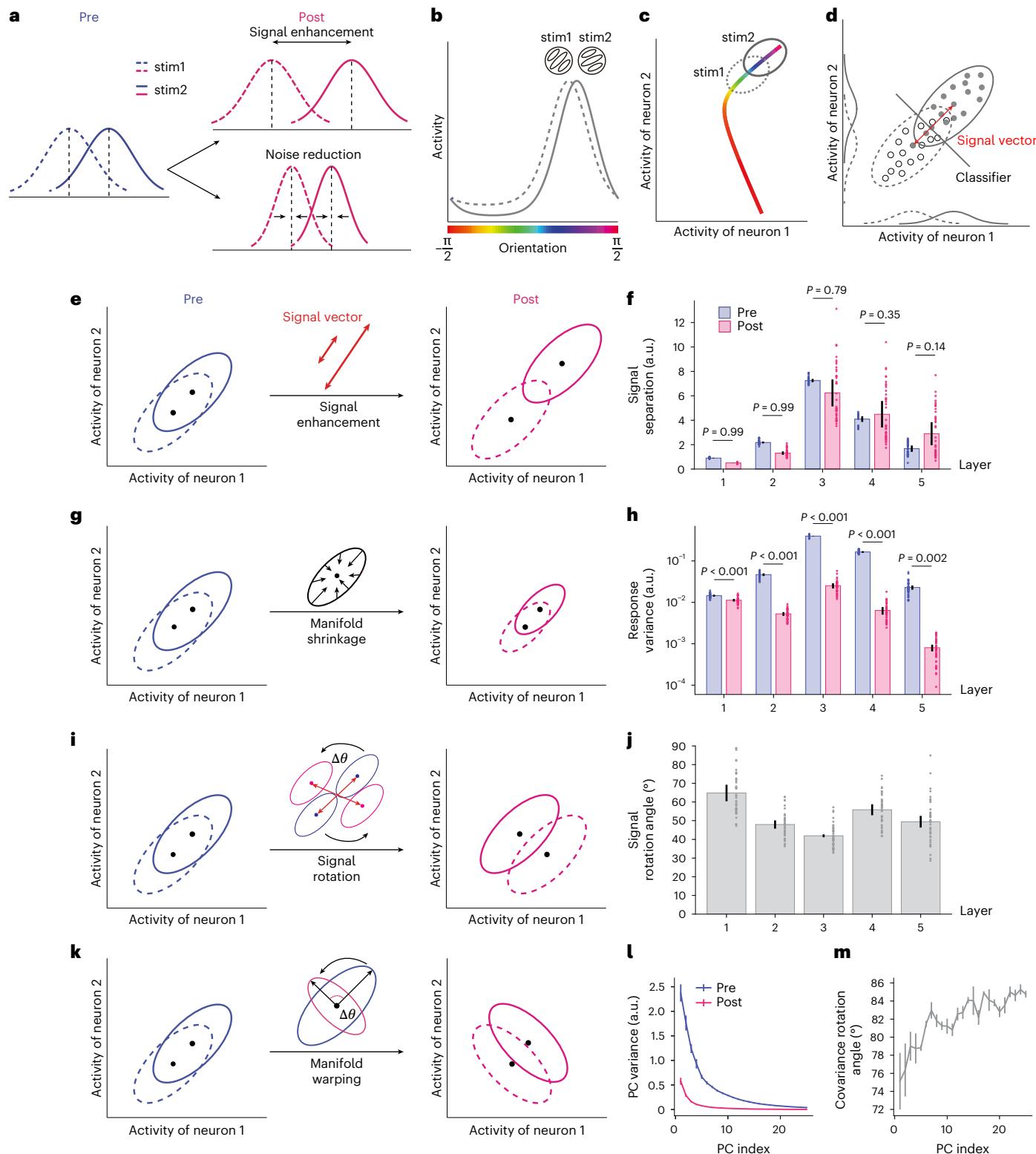
We identified the early visual areas (V1–V3), the motion-selective regions (V3A and hMT+) and the decision region (intraparietal sulcus, IPS) using independent functional localizer experiments (Fig. 6b). We estimated single-trial responses of voxels in these regions and then performed decoding analyses in these predefined regions, finding that motion training significantly enhanced decoding accuracy (Fig. 6c; V3A: one-sided paired *t*-test,  $t_{(21)} = -2.01$ ,  $P = 0.029$ , one-sided 95% CI  $-\infty$  to  $-3.7 \times 10^{-3}$ , Cohen’s  $d = -0.25$ ; hMT+: one-sided paired *t*-test,  $t_{(21)} = -1.95$ ,  $P = 0.032$ , one-sided 95% CI  $-\infty$  to  $-3.3 \times 10^{-3}$ , Cohen’s  $d = -0.50$ ) and aLFI (Fig. 6d) in areas V3A and hMT+ (V3A: one-sided paired *t*-test,  $t_{(21)} = -2.36$ ,  $P = 0.014$ , one-sided 95% CI  $-\infty$  to  $-5.0 \times 10^{-4}$ , Cohen’s  $d = -0.11$ ; hMT+: one-sided paired *t*-test,  $t_{(21)} = -1.99$ ,  $P = 0.030$ , one-sided 95% CI  $-\infty$  to  $-2.3 \times 10^{-4}$ , Cohen’s  $d = -0.47$ ), a result consistent with several human fMRI studies on motion VPL<sup>3,4,22</sup>.

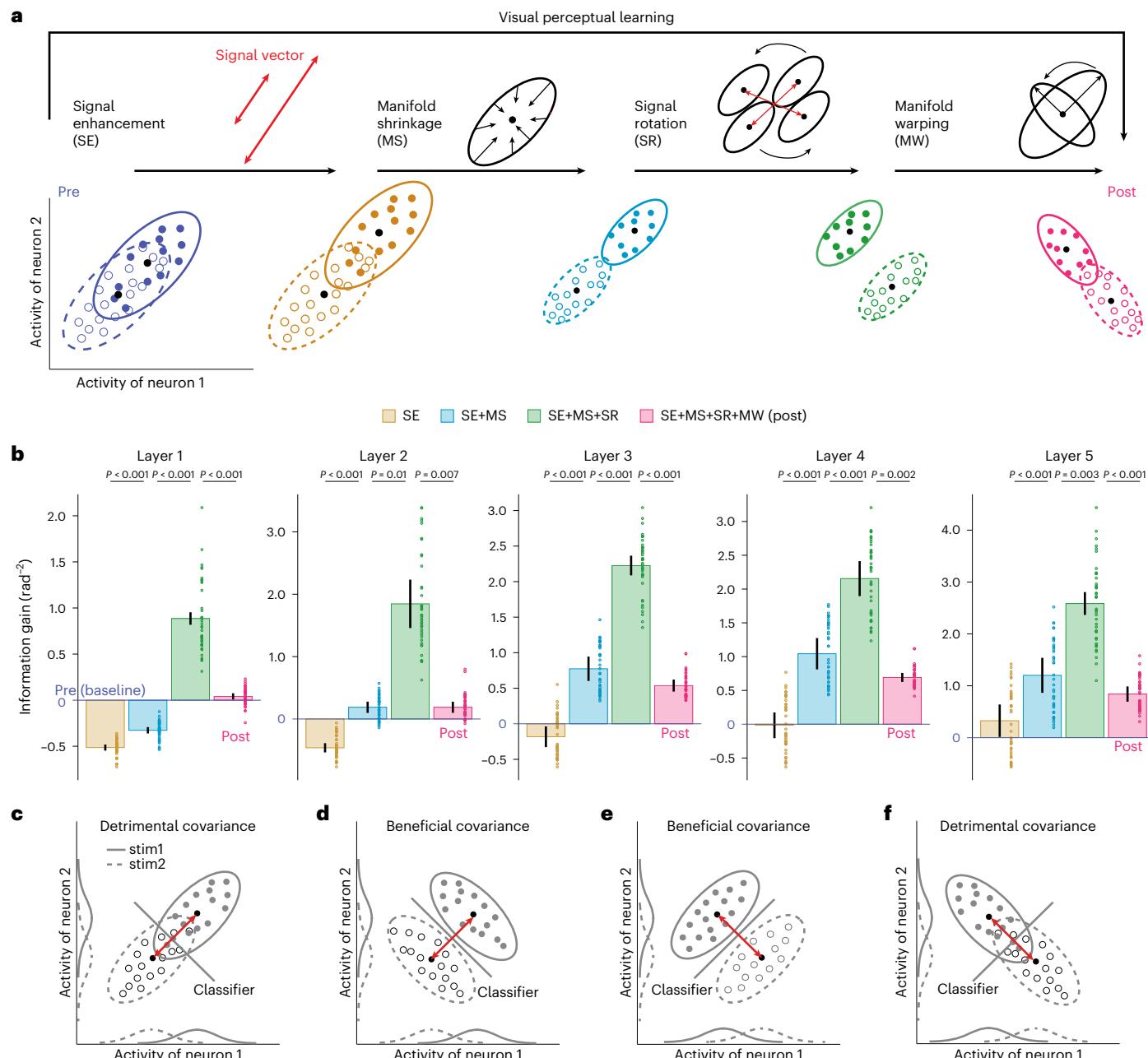
**Fig. 3 | Four possible mechanisms of VPL in neural populations.** **a**, To enhance sensory discriminability, the classical signal detection theory posits that signal enhancement predicts enlarged distances between two mean values while noise reduction predicts reduced variance of the two stimulus response distributions (stim1 and stim2). **b**, Stimulus orientation as a continuous stimulus variable can evoke high-dimensional population responses. **c**, If we continuously sweep the orientation value, the mean of population responses forms a closed-form ring in a high-dimensional neural space with dimensions equal to the number of units. The mean population responses to the two stimuli in a discrimination task are two points on the manifold. **d,e**, In realistic population responses, the trial-by-trial population responses to the two stimuli form two high-dimensional response distributions (that is, neural manifolds, **d**). The manifolds look elliptical rather than spherical due to pairwise noise correlations between units. In this high-dimensional neural space, the signal enhancement mechanism predicts an increased Euclidean distance (that is, signal separation, **e**) between two high-dimensional response distributions. **f**, However, no significant increase in signal

separation is observed in any of the five layers (signal separation decreases in the first two layers; one-sided paired *t*-test, all  $t_{(3)} > -1.27$ , all  $P > 0.146$ , all  $BF_{10} < 1.46$ ; see full statistical results in Supplementary Table 3). **g**, The manifold shrinkage mechanism predicts reduced variance of the two neural manifolds. **h**, This is observed in all five layers (one-sided paired *t*-test, all  $t_{(3)} > 8.39$ , all  $P < 0.002$ ; see full statistical results in Supplementary Table 4). **i**, The signal rotation mechanism predicts that the positions of the centroid (that is, mean) of the two manifolds are changed by training. **j**, The rotation angle ranges from approximately 50° to 70° in all five layers. **k**, The manifold warping mechanism predicts that training changes the shape of noise correlations. **l**, Indeed, training mostly reduces the variance of the high-variance principal components of the population responses. The principal components (showing only components that account for >99% of the total variance) are ranked from high to low variance. **m**, The directions of the principal components rotate from pre- to post-test. Data are presented as mean  $\pm$  s.e.m., with error bars and error shadings in **f–m** representing the s.e.m. across four ( $n = 4$ ) reference orientations.

We further investigated the coding principles in areas V3A and hMT+ and repeated the above analyses of DCNNs on fMRI data. Note that here we performed the same analyses on voxels instead of artificial neurons in DCNNs. Consistent with the predictions of the DCNNs, motion direction discrimination training in humans did not increase signal separation (Fig. 6e, V3A: one-sided paired *t*-test,  $t_{(21)} = 0.06, P = 0.526$ , one-sided 95%CI  $\sim 0$  to  $5.3 \times 10^{-2}$ , Cohen's  $d = 0.01$ , BF<sub>10</sub> 0.45; hMT+: one-sided paired *t*-test,  $t_{(21)} = 0.36, P = 0.639$ , one-sided 95%CI  $\sim 0$  to  $9.5 \times 10^{-2}$ , Cohen's  $d = 0.09$ , BF<sub>10</sub> 0.47) but markedly reduced

voxel response variance (Fig. 6f) in both areas (V3A: one-sided paired *t*-test,  $t_{(21)} = 2.87, P = 0.004$ , one-sided 95% CI  $2.9 \times 10^{-2}$  to  $\infty$ , Cohen's  $d = 0.16$ ; hMT+: one-sided paired *t*-test,  $t_{(21)} = 1.97, P = 0.031$ , one-sided 95% CI  $4.6 \times 10^{-3}$  to  $\infty$ , Cohen's  $d = 0.46$ ). Motion direction discrimination training also significantly reduced intervoxel correlations in hMT+ (Fig. 6g; one-sided paired *t*-test,  $t_{(21)} = 1.90, P = 0.035$ , one-sided 95% CI  $2.5 \times 10^{-3}$  to  $\infty$ , Cohen's  $d = 0.31$ ). The mechanism of signal rotation was also evident, as indicated by the average  $-55^\circ$  rotation of the signal vectors in both areas (Fig. 6h). In addition, training warped the





**Fig. 4 | Information decomposition in neural populations. a,** The effects of four mechanisms on population representations are decomposed into four distinct steps. **b,** The effects on information gain by sequentially adding each of the four mechanisms in each layer. For example, the increase in height from the brown to the blue bars indicates the positive contribution of manifold shrinkage to encoded stimulus information. Manifold shrinkage significantly increases the information (one-sided paired  $t$ -test, all  $t_{(3)} > 12.1$ , all  $P < 0.001$ ); signal rotation significantly increases the information (one-sided paired  $t$ -test, all  $t_{(3)} > 5.2$ , all  $P < 0.006$ ); manifold warping significantly decreases the information (one-sided paired  $t$ -test, all  $t_{(3)} > 4.3$ , all  $P < 0.01$ ). See full statistical results in Supplementary Table 5. **c-f,** Strong interaction effects between covariance and signal vector. For distributions with identical covariance (**c** and **e**; **d** and **f**), detrimental (**c** or **f**) or beneficial (**d** or **e**) effects on discriminability are possible, depending on the signal vector. Similarly, the effects of the signal vector also depend on its relative geometry to the axis of covariance. Data are presented as mean  $\pm$  s.e.m., with error bars and error shadings in **b** representing the s.e.m. across four ( $n = 4$ ) reference orientations.

paired  $t$ -test, all  $t_{(3)} > 4.3$ , all  $P < 0.01$ ). See full statistical results in Supplementary Table 5. **c-f,** Strong interaction effects between covariance and signal vector. For distributions with identical covariance (**c** and **e**; **d** and **f**), detrimental (**c** or **f**) or beneficial (**d** or **e**) effects on discriminability are possible, depending on the signal vector. Similarly, the effects of the signal vector also depend on its relative geometry to the axis of covariance. Data are presented as mean  $\pm$  s.e.m., with error bars and error shadings in **b** representing the s.e.m. across four ( $n = 4$ ) reference orientations.

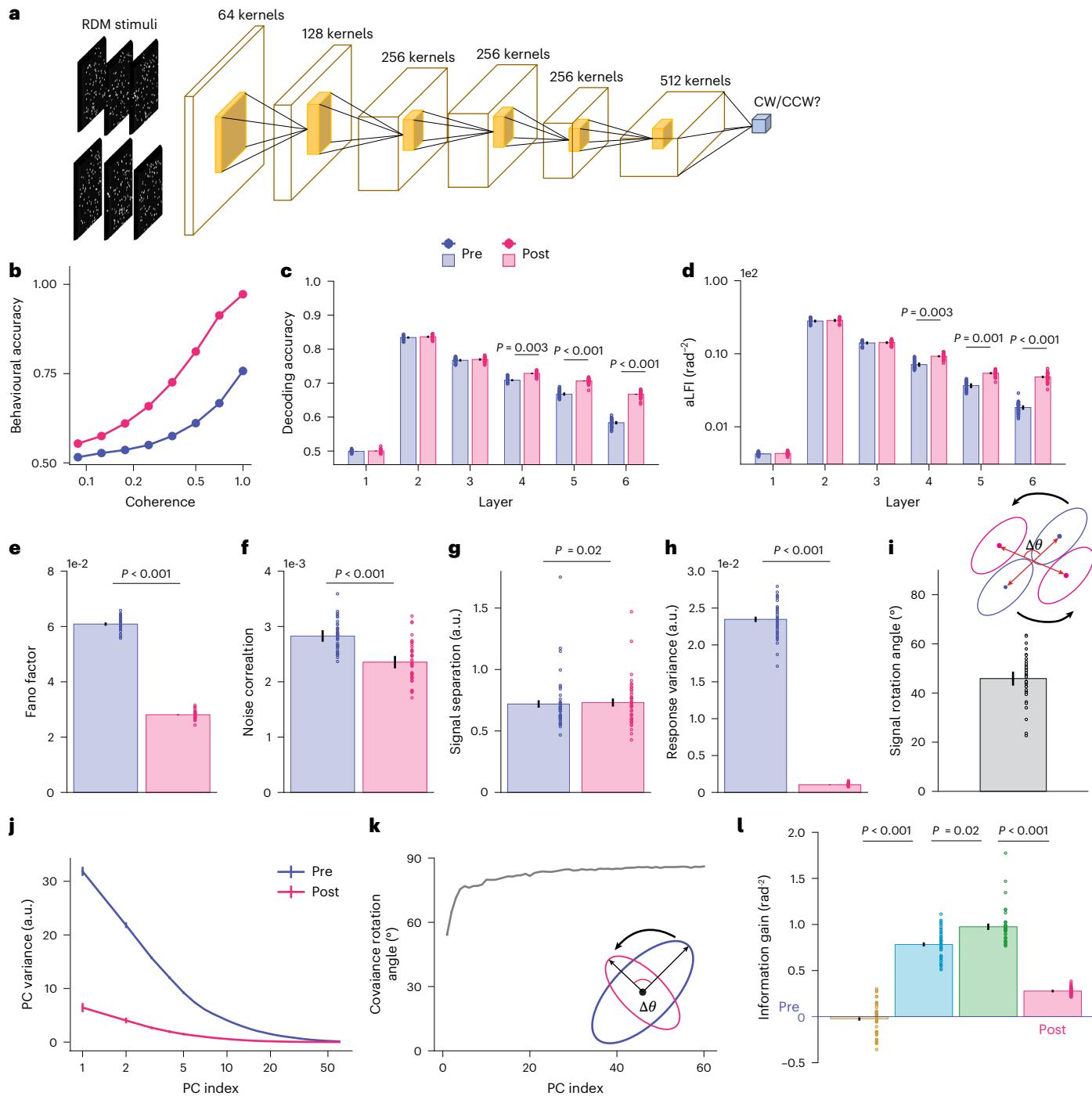
magnitude and direction of the covariance (Fig. 6*i,j*). Most importantly, the respective contributions of these four mechanisms in both brain regions were similar to the pattern in the DCNNs (Fig. 6*k*).

### Contrast discrimination learning in monkey V4

Voxel responses in fMRI studies reflect macroscopic brain activity that aggregates the responses of ~300,000–50,000 neurons<sup>38</sup>. It remains unclear whether the mechanisms we have discovered so far also exist at the local circuit level of single neurons or small clusters of neurons. To

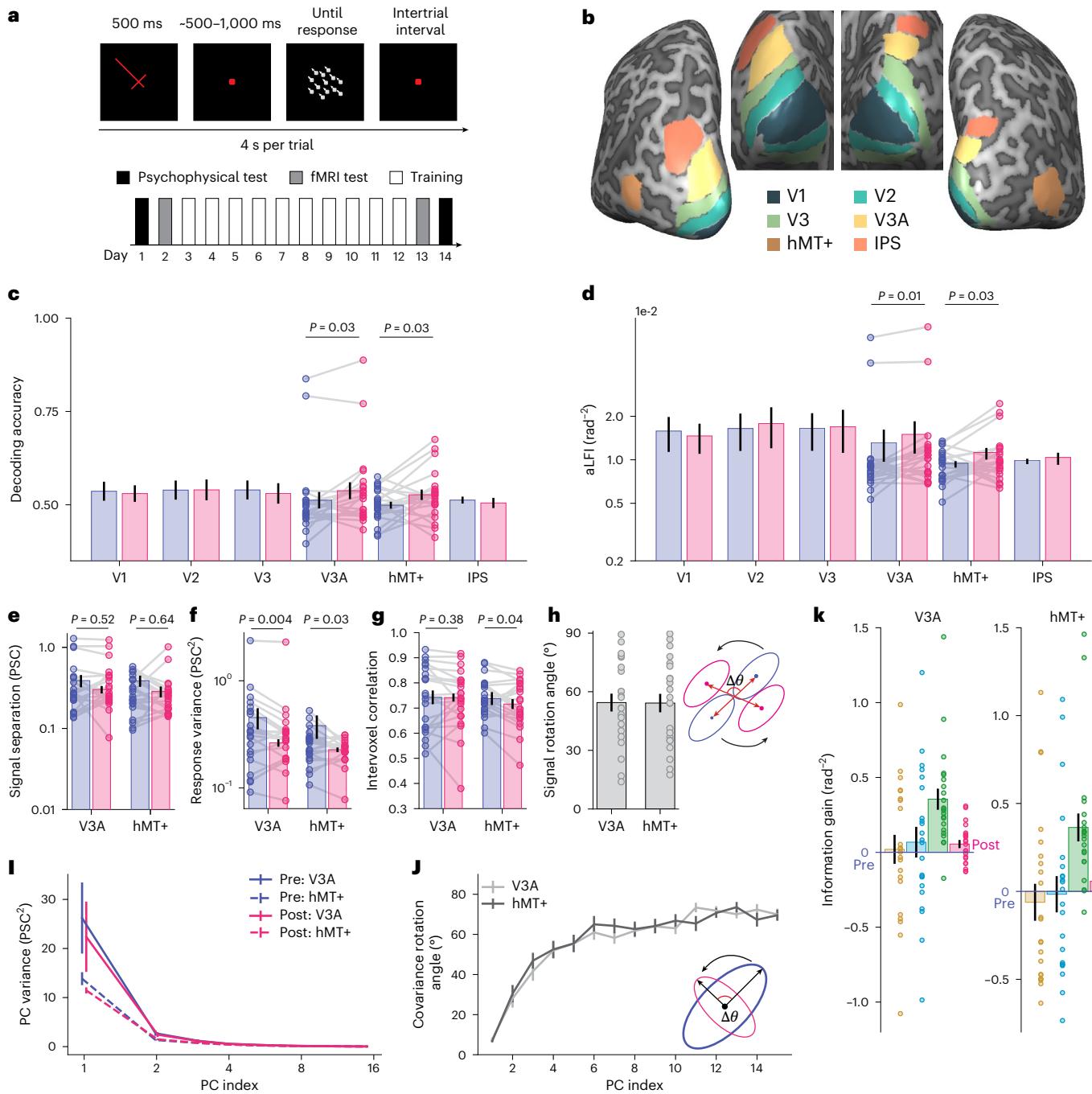
our knowledge, these predictions based on our neural network models have not been systematically tested using intracranial recording.

To further test our hypotheses on neuronal spiking activity, we analysed the population responses of V4 neurons in two monkeys (Fig. 7a) at the early stage and at the late stage of learning to perform a fine-contrast discrimination task (Fig. 7b, ref. 14). In this task, each monkey was presented sequentially with two identical Gabor patches with different contrast levels. The contrast of the reference (that is, the first) stimulus was always fixed at 30%, and the contrast of the target (that is,



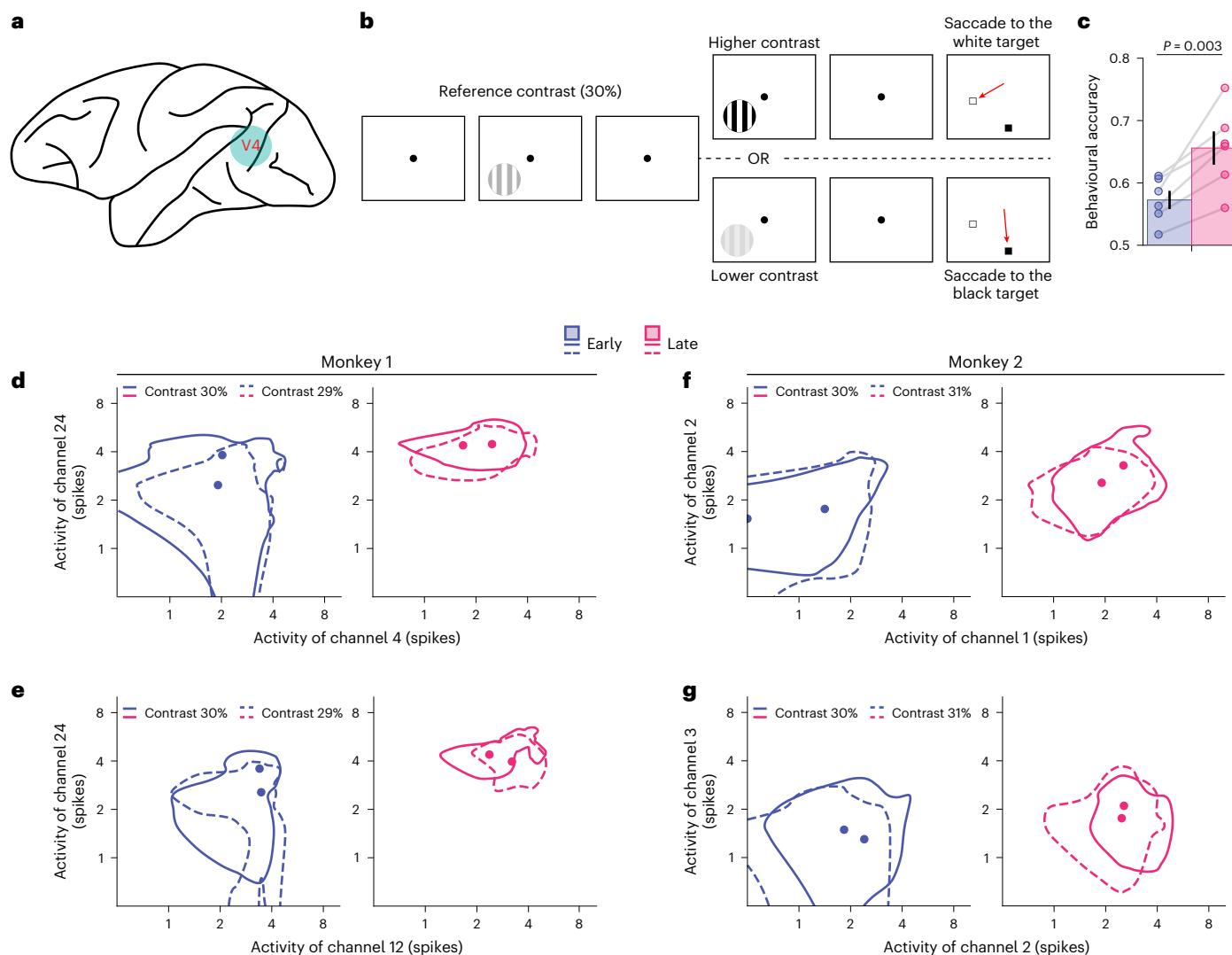
**Fig. 5 | DCNN modelling of motion VPL.** **a**, The DCNN of motion VPL uses 3D convolutions to process video stimuli. Here we simplify the four-dimensional feature maps in each convolutional layer and show them as 3D maps only for illustration purposes. **b–d**, Training improves DCNN direction discrimination performance (**b**), decoding accuracy (**c**; layers 4–6: one-sided paired  $t$ -test, all  $t_{(3)} < -7.14$ , all  $P < 0.028$ ; see full statistical results in Supplementary Table 6) and aLFI (**d**; layers 4–6: one-sided paired  $t$ -test, all  $t_{(3)} < -7.22$ , all  $P < 0.003$ ; see full statistical results in Supplementary Table 7). **e,f**, For single-unit analyses, motion direction discrimination training also reduces the Fano factor (**e**; one-sided paired  $t$ -test,  $t_{(3)} = 57.58$ ,  $P < 0.001$ , one-sided 95% CI  $3.1 \times 10^{-2}$  to  $\infty$ , Cohen's  $d = 38.75$ ) and noise correlation (**f**; one-sided paired  $t$ -test,  $t_{(3)} = 42.84$ ,  $P < 0.001$ , one-sided 95% CI  $14.4 \times 10^{-4}$  to  $\infty$ , Cohen's  $d = 2.19$ ) in layer 6. **g,i**, Similar to orientation discrimination training, motion direction discrimination training does not significantly enhance signal separation (**g**; one-sided paired  $t$ -test,

$t_{(3)} = -0.98$ ,  $P = 0.198$ , one-sided 95% CI  $-\infty$  to  $1.7 \times 10^{-2}$ , Cohen's  $d = -0.19$ ,  $BF_{10} = 1.22$ ) but rotates the position of the two distributions in layer 6 (**i**). **h**, Importantly, training clearly reduces the response variance in layer 6 (one-sided paired  $t$ -test,  $t_{(3)} = 59.05$ ,  $P < 0.001$ , one-sided 95% CI  $2.2 \times 10^{-2}$  to  $\infty$ , Cohen's  $d = 43.89$ ). **j,k**, Specifically, training reduces the variance of the high-variance PCs (**j**) and rotates the directions of all PCs (**k**), indicating a significant effect of manifold warping in layer 6. **l**, The pattern of information gain associated with the four possible mechanisms is consistent with that of orientation discrimination training (one-sided paired  $t$ -test,  $t_{(3)} = 76.0$ ,  $P < 0.001$  for manifold shrinkage,  $t_{(3)} = 3.80$ ,  $P = 0.02$  for signal rotation,  $t_{(3)} = 17.7$ ,  $P < 0.001$  for manifold warping). See results for all six layers in Supplementary Fig. 3. Data are presented as mean  $\pm$  s.e.m., with error bars and error shadings in **c–l** representing the s.e.m. across four ( $n = 4$ ) reference directions. Note that some error bars are very small and barely visible.



**Fig. 6 | Motion VPL induces neural geometry changes in the human brain.** **a,b**, Trial diagram and training paradigm (a), and ROIs in a typical subject (b). **c,d**, Motion direction discrimination training in humans significantly improves decoding accuracy (c; for V3A: one-sided paired  $t$ -test,  $t_{(21)} = -2.01, P = 0.029$ , one-sided 95% CI  $\leftarrow \infty$  to  $-3.7 \times 10^{-3}$ , Cohen's  $d = -0.25$ ; for hMT+: one-sided paired  $t$ -test,  $t_{(21)} = -1.95, P = 0.032$ , one-sided 95% CI  $\leftarrow \infty$  to  $-3.3 \times 10^{-3}$ , Cohen's  $d = -0.50$ ) and aLFI (d; for V3A: one-sided paired  $t$ -test,  $t_{(21)} = -2.36, P = 0.014$ , one-sided 95% CI  $\leftarrow \infty$  to  $-5.0 \times 10^{-4}$ , Cohen's  $d = -0.11$ ; for hMT+: one-sided paired  $t$ -test,  $t_{(21)} = -1.99, P = 0.030$ , one-sided 95% CI  $\leftarrow \infty$  to  $-2.3 \times 10^{-4}$ , Cohen's  $d = -0.47$ ) in areas V3A and hMT+, a finding consistent with several existing fMRI studies of motion VPL. Note that the four data points in V3A appear as outliers in c and d, but the results still hold if these data points are removed. **e–g**, Motion direction discrimination training does not significantly change signal separation in V3A and hMT+ (e; for V3A: one-sided paired  $t$ -test,  $t_{(21)} = 0.06, P = 0.526$ , one-sided 95% CI  $\leftarrow \infty$  to  $5.3 \times 10^{-2}$ , Cohen's  $d = 0.01$ ,  $\text{BF}_{10} 0.45$ ; for hMT+: one-sided paired  $t$ -test,  $t_{(21)} = 0.36, P = 0.639$ , one-sided 95% CI  $\leftarrow \infty$  to  $9.5 \times 10^{-2}$ , Cohen's  $d = 0.09$ ,  $\text{BF}_{10} 0.47$ ) but reduces voxel response variance in V3A and hMT+ (f; for V3A: one-sided paired  $t$ -test,  $t_{(21)} = 2.87, P = 0.004$ , one-sided 95% CI  $12.9 \times 10^{-2}$  to  $\infty$ , Cohen's  $d = 0.16$ ; for hMT+: one-sided paired  $t$ -test,  $t_{(21)} = 1.97, P = 0.031$ , one-sided 95% CI  $4.6 \times 10^{-3}$  to  $\infty$ , Cohen's  $d = 0.46$ ) and intervoxel noise correlations in hMT+ (g; one-sided paired  $t$ -test,  $t_{(21)} = 1.90, P = 0.035$ , one-sided 95% CI  $2.5 \times 10^{-3}$  to  $\infty$ , Cohen's  $d = 0.31$ ). **h–j**, Similar to the motion DCNNs, motion direction discrimination training in humans also rotates stimulus distributions (h), reduces the variance of high-variance PCs (i) and warps the covariance directions (j). **k**, The patterns of information gain associated with the four mechanisms are consistent with those in the DCNNs of both orientation and motion VPL. The unit PSC represents percent signal change of BOLD signals. Individual data points represent the human subjects. Data are presented as mean  $\pm$  s.e.m., with error bars in all panels representing the s.e.m. across subjects ( $n = 22$ ). \* $P < 0.05$ , \*\* $P < 0.01$ .

$t$ -test,  $t_{(21)} = 0.36, P = 0.639$ , one-sided 95% CI  $\leftarrow \infty$  to  $9.5 \times 10^{-2}$ , Cohen's  $d = 0.09$ ,  $\text{BF}_{10} 0.47$ ) but reduces voxel response variance in V3A and hMT+ (f; for V3A: one-sided paired  $t$ -test,  $t_{(21)} = 2.87, P = 0.004$ , one-sided 95% CI  $12.9 \times 10^{-2}$  to  $\infty$ , Cohen's  $d = 0.16$ ; for hMT+: one-sided paired  $t$ -test,  $t_{(21)} = 1.97, P = 0.031$ , one-sided 95% CI  $4.6 \times 10^{-3}$  to  $\infty$ , Cohen's  $d = 0.46$ ) and intervoxel noise correlations in hMT+ (g; one-sided paired  $t$ -test,  $t_{(21)} = 1.90, P = 0.035$ , one-sided 95% CI  $2.5 \times 10^{-3}$  to  $\infty$ , Cohen's  $d = 0.31$ ). **h–j**, Similar to the motion DCNNs, motion direction discrimination training in humans also rotates stimulus distributions (h), reduces the variance of high-variance PCs (i) and warps the covariance directions (j). **k**, The patterns of information gain associated with the four mechanisms are consistent with those in the DCNNs of both orientation and motion VPL. The unit PSC represents percent signal change of BOLD signals. Individual data points represent the human subjects. Data are presented as mean  $\pm$  s.e.m., with error bars in all panels representing the s.e.m. across subjects ( $n = 22$ ). \* $P < 0.05$ , \*\* $P < 0.01$ .



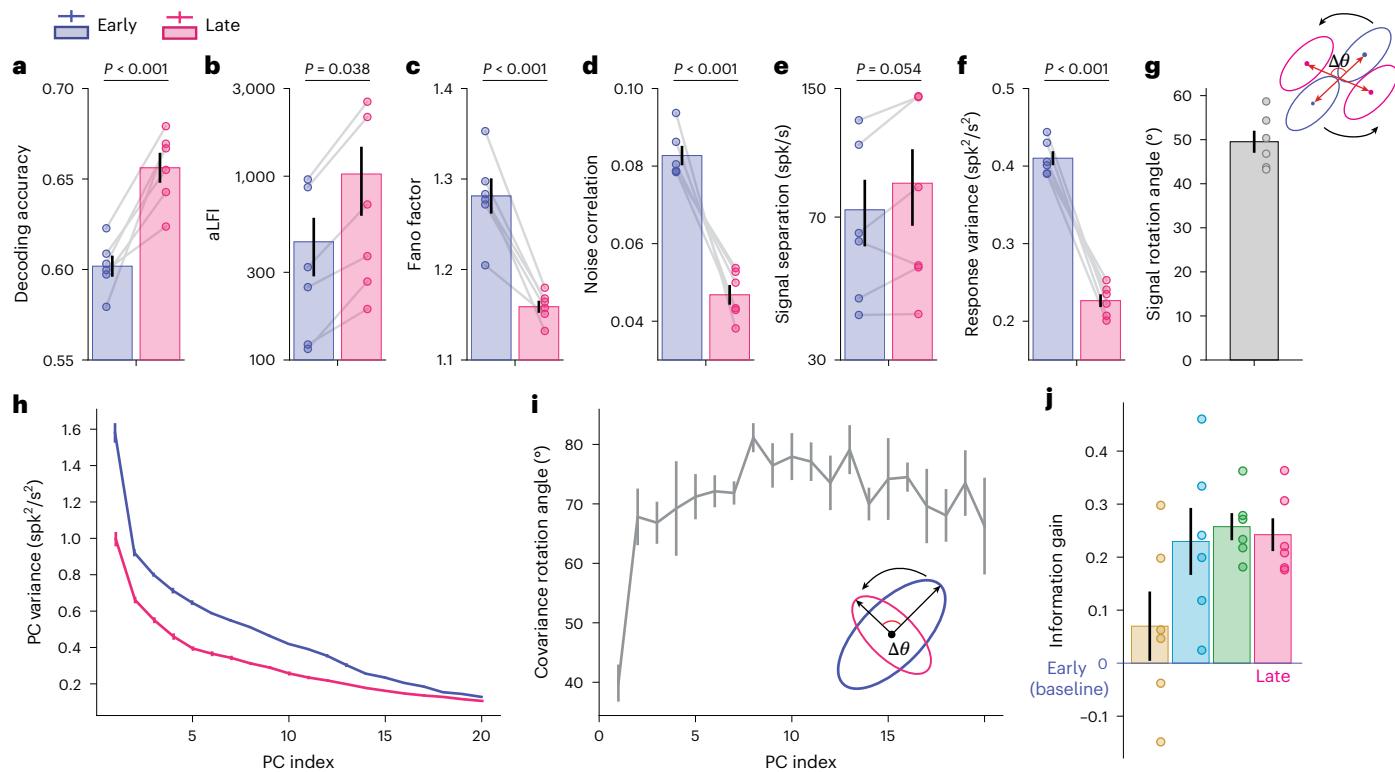
**Fig. 7 | Single-unit analyses of contrast discrimination learning in monkey V4.** **a,b**, We analysed population responses in area V4 (**a**) of two monkeys while they were trained on a fine contrast discrimination task (**b**). The first four and last four training sessions were grouped as pre- and the post-test conditions, respectively. Contrast discrimination training significantly improved behavioural performance from the early to late stage of training (**c**; one-sided paired *t*-test,  $t_{(5)} = -4.61, P = 0.003$ , Cohen's  $d = -1.57$ ). **c**, All individual data points represent the six target contrast conditions (27%, 28%, 29%, 31%, 32% and 33%; the reference

contrast is 30%). Each point is averaged over the two monkeys. See plots for individual monkeys in Supplementary Fig. 4. Data are presented as mean  $\pm$  s.e.m., with error bars indicate the s.e.m. across the six conditions ( $n = 6$ ). **d–g**, The full width at half maximum of the response distributions of four pairs of channels at pre- and post-test (**d** and **e** for monkey 1 and **f** and **g** for monkey 2). The solid lines represent 30% reference contrast, and the dashed lines represent 29% and 31% target contrast in monkey 1 and monkey 2, respectively. These results show that learning systematically changes the geometries of the multivariate responses.

the second) stimuli varied systematically near the reference contrast (that is, 27%, 28%, 29%, 31%, 32% and 33%). This contrast discrimination training significantly improved behavioural performance (Fig. 7c; one-sided paired *t*-test,  $t_{(5)} = -4.61, P = 0.003$ , one-sided 95% CI  $\sim$  to  $-4.7 \times 10^{-2}$ , Cohen's  $d = -1.57$ ). Most importantly, responses of multiple channels were continuously recorded via chronically implanted electrodes in area V4 (29 and 20 channels for monkeys 1 and 2, respectively) throughout training (21 and 23 training sessions for the two monkeys, respectively). This continuous multiunit recording is the key to disentangling population-level changes associated with VPL.

We used the above analyses (previously applied to DCNNs and human fMRI data) and applied them to the monkey V4 responses, and again found highly consistent results (see results of each monkey in Supplementary Fig. 4). First, contrast discrimination training significantly improved stimulus information at the population level (Fig. 8a,b; decoding accuracy: one-sided paired *t*-test,  $t_{(5)} = -6.03$ ,

$P < 0.001$ , one-sided 95% CI  $\sim$  to  $-3.6 \times 10^{-2}$ , Cohen's  $d = -3.10$ ; aLFI: one-sided paired *t*-test,  $t_{(5)} = -2.21, P = 0.039$ , one-sided 95% CI  $\sim$  to  $-52$ , Cohen's  $d = -0.76$ ). Second, at the individual level, contrast discrimination training also significantly reduced Fano factors (Fig. 8c; one-sided paired *t*-test,  $t_{(5)} = 7.28, P < 0.001$ , one-sided 95% CI  $18.8 \times 10^{-2}$  to  $\infty$ , Cohen's  $d = 3.43$ ) and noise correlations (Fig. 8d; one-sided paired *t*-test,  $t_{(5)} = 7.46, P < 0.001$ , one-sided 95% CI  $2.6 \times 10^{-2}$  to  $\infty$ , Cohen's  $d = 5.80$ ), consistent with several existing findings. Interestingly, while the trial-by-trial variance was significantly reduced after training (Fig. 8f; one-sided paired *t*-test,  $t_{(5)} = 13.24, P < 0.001$ , one-sided 95% CI  $1.6 \times 10^{-1}$  to  $\infty$ , Cohen's  $d = 8.70$ ), no apparent change in signal separation was observed (Fig. 8e; one-sided paired *t*-test,  $t_{(5)} = -1.957, P = 0.054$ , one-sided 95% CI  $-3.7 \times 10^{-1}$  to  $\infty$ , Cohen's  $d = -0.30$ ,  $BF_{10} = 2.41$ ), suggesting the predominant role of manifold shrinkage. Importantly, we again observed evidence for signal rotation (Fig. 8g) and manifold warping (Fig. 8h,i). The stepwise information analyses also qualitatively replicated the relative contributions of



**Fig. 8 | Population activity analyses of contrast discrimination learning in monkey V4.** **a, b,** Contrast discrimination training significantly enhanced stimulus information at the population level (for decoding accuracy: one-sided paired  $t$ -test,  $t_{(5)} = -6.03, P < 0.001$ , one-sided 95% CI  $\sim 0$  to  $-3.6 \times 10^{-2}$ , Cohen's  $d = -3.10$  (**a**); for aLFI: one-sided paired  $t$ -test,  $t_{(5)} = -2.21, P = 0.039$ , one-sided 95% CI  $\sim 0$  to  $-52$ , Cohen's  $d = -0.76$  (**b**)). **c–f,** Consistent with VPL in the DCNNs and the human brain, training monkeys on a contrast discrimination task reduced Fano factors (**c**; one-sided paired  $t$ -test,  $t_{(5)} = 7.28, P < 0.001$ , one-sided 95% CI  $8.8 \times 10^{-2}$  to  $\infty$ , Cohen's  $d = 3.43$ ), noise correlations (**d**; one-sided paired  $t$ -test,  $t_{(5)} = 7.46, P < 0.001$ , one-sided 95% CI  $1.26 \times 10^{-2}$  to  $\infty$ , Cohen's  $d = 5.80$ ) and response variance (**f**; one-sided paired  $t$ -test,  $t_{(5)} = 13.24, P < 0.001$ , one-sided 95% CI

$1.6 \times 10^{-1}$  to  $\infty$ , Cohen's  $d = 8.70$ ) but had no significant effect on signal separation (**e**; one-sided paired  $t$ -test,  $t_{(5)} = -1.957, P = 0.054$ , one-sided 95% CI  $-3.7 \times 10^{-1}$  to  $\infty$ , Cohen's  $d = -0.30$ , BF<sub>10</sub> 2.41). **g–i,** We also found evidence for signal rotation (**g**) and manifold warping (**h** for PC variance and **i** for PC rotation). **j,** The stepwise information analyses also show the similar pattern of the four mechanisms. The unit 'spk/s' indicates the number of spikes per second (that is, firing rate). We calculate aLFI and information gain using stimulus contrast as decimal values (that is, 0.29), so they have arbitrary units. Each point is averaged over the two monkeys. See plots for individual monkeys in Supplementary Fig. 4. Data are presented as mean  $\pm$  s.e.m., with error bars indicating the s.e.m. across the six conditions ( $n = 6$ ).

the four mechanisms to the total stimulus information encoded in the population (Fig. 8j).

## Discussion

It has been controversial whether single-unit properties such as sharpened tuning curves<sup>9,10</sup> or reduction of noise correlations<sup>11,12</sup> contribute to VPL. Our information-theoretic analysis on neural geometry suggested that, although these changes were indeed observed, they did not contribute significantly to the improved population representations associated with VPL. Rather, we found that the totally overlooked mechanism—the response variance of individual units (that is, manifold shrinkage)—is the primary contributor to the improved population representations associated with VPL. These results were further tested on DCNNs, human fMRI data and monkey neurophysiological data associated with different VPL tasks and brain regions.

Given the pronounced changes in tuning curves and noise correlations observed after training, why do they not contribute to VPL? Conventional approaches treat changes in tuning curves and in noise correlations as two independent factors mediating VPL. However, according to the neural geometry approach, the effects of tuning curve changes can be decomposed into two parts: signal enhancement independent of noise correlations and signal rotation interacting with noise correlations (equation (4) in Methods). We observed minimal contributions of signal enhancement to population

representations. Although we observed the phenomena of signal rotation and manifold warping, their respective contributions appeared significant but their overall joint effects were minimal because their respective effects can cancel each other out (Supplementary Fig. 5).

Our finding that manifold shrinkage is the primary contributor to improved population representations is of unique significance in constraining the model of VPL. We note that the goal of perceptual learning is to produce more discriminable population representations such that downstream decision units can easily read out sensory information. However, deciphering the underlying format of discriminable representations is non-trivial because discriminable representations can be achieved by any or combinations of four possible mechanisms. The key contribution of our work lies in the systematic quantification of the four mechanisms. In manifold shrinkage, the total variance of the high-dimensional distributions is scaled down (that is,  $\lambda$  in equation (4) is reduced). In other words, the two stimulus distributions simply shrink to a smaller size (Fig. 4). Note that manifold shrinkage is independent of any tuning changes and noise correlation changes. We also emphasize that manifold shrinkage and manifold warping are two different mechanisms. In our approach, manifold warping redistributes the variance of the high-dimensional distributions in different directions (that is,  $\lambda_i$  and  $\xi_i$  in equation (4) are changed) but, unlike manifold shrinkage, the total amount of variance remains unchanged. Thus, the shape of the two stimulus distributions is significantly warped. We thus emphasize manifold shrinkage as a marker of global

population responses to differentiate it from trial-by-trial variability changes in single units.

Our neural geometry approach is consistent with the earlier applications of high-dimensional signal detection theory (MSDT) in psychophysics and systems neuroscience<sup>39,40</sup>. MSDT is a powerful tool for quantifying the discriminability of population representations. Although we also focus on the discriminability of population representations during the learning process, we extend this framework by conceptualizing MSDT as neural geometries to bridge changes in individual-level responses and changes in population representations. Specifically, changes in individual responses (for example, changes in tuning curves and/or noise correlations) are characterized as geometric transformations (for example, signal enhancement and manifold warping) of neural manifolds. These geometric transformations elucidate how discriminability in high-dimensional response distributions is enhanced. This geometric perspective enables experimentally testable predictions about learning effects of on neural manifolds, offering insights to adjudicate previous theories of VPL.

This high-dimensional geometric approach has been used in topics such as classification<sup>41</sup>, attention<sup>42</sup> and neural coding<sup>43</sup>. The geometric similarities also predict perceptual similarities in humans<sup>44</sup>. A recent study<sup>45</sup> found that such coordinates are not arbitrary, but privileged. The high-dimensional representational axes are highly consistent across different humans and even across different DCNNs. These representational axes lead to better readout or generalization abilities.

Our work unifies several important existing findings of VPL. First, it has long been hypothesized that noise reduction is an important mechanism of VPL<sup>6,7,21</sup>, but the exact underlying neural mechanisms remain elusive. Our work demonstrates that at least manifold shrinkage due to reduced trial-by-trial response variability is a viable mechanism to support noise reduction. Second, Bejjanki et al.<sup>46</sup> built a biological neural network and, similar to our task, simulated the effects of orientation VPL on Gabor stimuli with different levels of image noise. The results showed that changes in orientation tuning curve have only modest effects on psychophysical TvN functions. Using a different network architecture (pretrained artificial DCNNs), our study replicated the finding of sharpened orientation-selective tuning curves reported and also showed that the effects of such tuning changes are modest. Our modelling here suggests that sharpened tuning curves do not necessarily lead to improved population codes, given that other aspects of population responses are also changed by learning. Third, most existing human imaging studies and single-unit studies on VPL have focused only on changes in population representations<sup>3,4,22</sup> or changes in individual neurons<sup>9,10</sup>, respectively. Previous studies attempted to address the relationship between the two levels by projecting high-dimensional neural manifolds onto a one-dimensional optimal decision plane<sup>13,22</sup>. However, we argue that this approach is inadequate (see analytical derivations in Supplementary Note 2) and we should explicitly disentangle and quantify the effects of individual factors (see additional analysis in Supplementary Fig. 6).

It is noteworthy that our approach is based on the assumption that VPL is associated with changes in neuronal populations. However, we do not dismiss all neuron-level accounts for VPL. For example, VPL could be conceptualized as a search in neuronal space for the most informative neurons for the trained task. These neurons are not necessarily the ones most responsive to the trained stimuli or those that represent them most efficiently. For example, post-adaptation orientation discrimination in expert subjects has been shown to involve learning that the most informative channel/filter for discrimination is rotated about 10–20° away from the observed stimulus<sup>47</sup>. Similar results were observed in monkey neurons during training of VPL of orientation discrimination<sup>9</sup>. The specific rotation magnitude may depend on the tuning curves and noise properties of the neurons.

Our study still has several limitations that could be addressed by future studies. First, although DCNN has recently emerged as a promising computational framework for modelling, there still exist clear differences between DCNNs and biological visual systems. Our models here are all feedforward architectures and lack the component of top-down modulation. Top-down modulation is an important aspect of supervised training<sup>48</sup> and particularly useful for considering within-trial neural dynamics<sup>49</sup>. Second, VPL can be achieved by unsupervised training<sup>50</sup> or even pure mental imagery<sup>51</sup>. These learning regimes cannot be explained by current models. Third, this study examines only how VPL improves population codes of trained stimuli. It remains unclear how learning effects generalize to other untrained stimuli, which is recently proposed as a key question in VPL<sup>52</sup>. Fourth, it remains unclear the perceptual consequences predicted by our neural geometry approach, especially by each mechanism. To address this, we conducted thorough simulations of neural geometric changes and derived their predictions on perceptual detection and perceptual estimation tasks (Supplementary Note 3 and Supplementary Fig. 7), which could be further tested in future studies. Our framework also provides a theoretical foundation to understand neural underpinnings of generalization in future studies.

## Methods

### DCNN modelling of orientation VPL

**Stimuli.** The network was trained to discriminate whether a target stimulus was tilted 1° clockwise or counterclockwise relative to a reference stimulus. All reference stimuli in the orientation discrimination task were Gabor patterns (227 × 227 pixels; spatial frequency, 40 pixels per cycle; standard deviation of the Gaussian spatial envelope, 50 pixels). The stimuli were varied in contrast (0.1 to 1.0 in 0.1 increments) and image noise level (eight levels: 0.005, 1, 5, 10, 15, 30, 50 and 75). Similar to existing psychophysical studies<sup>53</sup>, the image noise level is defined as the fraction of pixels randomly selected and replaced by Gaussian noise with a standard deviation of 15 gray level units. To mimic intrinsic sensory noise, we also added Gaussian white noise (standard deviation 10) to each stimulus<sup>19</sup>. To match the spatial frequency of noise and signal, the size of the replaced pixels was set to be 8 × 8. Four reference orientations (35°, 55°, 125° and 145°) were used, and we trained ten DCNNs (ten different random seeds, see below) for each of the four reference orientations. This yields 40 DCNNs models of VPL.

**Neural networks and training.** A DCNN<sup>20</sup> was used to simulate the orientation VPL. We retained the first five convolutional layers of the pretrained AlexNet and replaced its three fully connected layers with a single linear fully connected layer for perceptual choice. The network was configured in a Siamese fashion to perform the two-alternative forced-choice task: the same network was fed with both the target and the reference stimuli, producing two scalar outputs,  $h_t$  and  $h_r$ , respectively. The network then made the final decision with a probability  $p$  (classification confidence) calculated by the sigmoid function

$$p = \frac{e^{h_t - h_r}}{1 + e^{h_t - h_r}}. \quad (1)$$

The entire training procedure consisted of two distinct phases: the pretraining phase and the VPL phase. In the pretraining phase, the network was trained on full-contrast noiseless stimulus pairs to understand the task and to establish the pre-test baseline. In the VPL phase, the network was trained on stimulus pairs across all contrasts (ten levels) and noise levels (eight levels). The network was trained for 5,000 epochs in the pretraining phase and 500 epochs in the VPL phase using the stochastic gradient descent learning algorithm. The learning rate and the momentum were set to  $10^{-5}$  and 0.9, respectively. The parameters were updated to minimize the cross-entropy loss between the network outputs and the true stimulus labels. The initial

parameters in the fully connected layer were set to zero, as in ref. 19, while those in the convolutional layers were taken directly from a pretrained AlexNet available at [http://dl.affe.berkeleyvision.org/bvlc\\_AlexNet.caffemodel](http://dl.affe.berkeleyvision.org/bvlc_AlexNet.caffemodel). We trained one model for each of the four reference orientations, and the entire procedure was repeated ten times for each reference orientation to control for randomness. All model and training procedures were implemented using Python 3.10.9 conda environment, including pytorch 1.13.1, scikit-learn 1.2.0. Details of the full conda environment are provided via GitHub at [https://github.com/Yu-AngCheng/neural\\_geometry\\_VPL](https://github.com/Yu-AngCheng/neural_geometry_VPL).

**Behavioural and neural changes.** For each reference orientation, we used the stimuli with the same orientations in pre-/post-tests and in training phase. The only difference is that stimulus images were randomly generated in each trial. We derived the behavioural psychometric curves of the network before and after the VPL phase defined above. Specifically, the behavioural performance of the network was evaluated by measuring its classification confidence (equation (1)) at all 80 conditions (10 contrast levels  $\times$  8 noise levels) with 1,000 trials in each condition. The classification confidence of all 1,000 trials was averaged (Fig. 1c,d). The behavioural TvN curves (Fig. 1f) of the model were further derived for comparison with human psychophysical results. Specifically, for each noise level, a contrast threshold was obtained by interpolating accuracy–contrast psychometric curves at the accuracies of 55% and 70% for pre-test and post-test respectively.

To quantify the activity of artificial neurons, in each trial, the firing rate of each artificial neuron was measured as the output of local response normalization or rectified linear unit (ReLU) layers, averaged over all locations. All measurements were obtained by simulating 1,000 trials for better estimation. To ensure that units were truly driven by the stimuli, only units with a mean firing rate greater than 0.001 before and after training were included in the analyses<sup>19</sup>. To perform population decoding analyses, we trained a linear classifier on the firing rates of the artificial neurons to discriminate the target and the reference stimuli. The classifier was trained on half of the 1,000 simulated trials, while the other half served as the test dataset.

To characterize the response properties of individual units, we measured orientation-selective tuning curves by sweeping the orientation of high-contrast stimuli from 0° to 180°. The tuning curves were derived by averaging 100 simulated trials for each orientation. The resulting tuning curves were then smoothed with a 10° Gaussian kernel. To control the heterogenous response range across units, we then normalized the tuning curves of each unit by its maximum response and averaged the tuning curves across units to obtain the group-level tuning curves. The group-level tuning curves were then fitted with a Gaussian function and rescaled to -0–1 for better comparison.

To calculate the Fano factor of each unit, we simulated 1,000 trials for each reference orientation. The Fano factor of each artificial neuron is defined as the ratio of the variance of the firing rate to its mean. Similarly, noise correlations between artificial neurons were calculated as the correlations between unit firing rates over the 1,000 simulated trials for each reference orientation. We took the median of the Fano factor across units in each layer to generate the data plot (Fig. 2i). We took the median of the lower triangle of the noise correlation matrix in each layer to generate the data plot (Fig. 2j). The error bars in Fig. 2i,j represent the standard errors across four reference orientations.

**Linear Fisher information analyses.** To understand how neural activation contributes to behavioural improvements, we applied linear Fisher information analysis to population responses. We considered the firing rates of the same groups of units under the reference and the target stimulus conditions as two distributions in a high-dimensional neural space. We refer to the signal vector as the vector connecting the mean of the two distributions. A signal vector is calculated as the

difference between the mean firing rates of units to two stimuli. The signal separation is referred to as the modulus length of the signal vector, and the angle of the signal vectors before and after training is referred to as the signal rotation angle.

To measure how much information was contained in a layer per unit, we calculated the aLFI as follows:

$$aLFI = \frac{1}{n} \cdot \frac{df^T \bar{\Sigma}^{-1} df}{\Delta\theta^2} \quad (2)$$

$$\begin{aligned} \bar{\Sigma} &= \frac{\Sigma_1 + \Sigma_2}{2} \\ &= \frac{V_1^T C_1 V_1 + V_2^T C_2 V_2}{2}, \end{aligned} \quad (3)$$

where  $n$  is the number of units in a layer,  $\Delta\theta$  is the separation between the target stimulus and the reference stimulus (that is, 1°),  $df$  is the signal vector,  $\bar{\Sigma}$  is the mean of the covariance matrices (that is,  $\Sigma_1$  and  $\Sigma_2$ ) of units responding to the two stimuli,  $V$  is a diagonal matrix with the variance of the units as the diagonal terms, and  $C$  is the correlation matrix of the population with all diagonal elements equal to 1.

To further elaborate on the potential mechanisms of the improved LFI, we performed an eigendecomposition on the covariance matrix  $\bar{\Sigma}$ , where we obtained  $\lambda_i$ , the eigenvalue of  $\bar{\Sigma}$ , and  $\xi_i$ , its corresponding normalized eigenvector. The aLFI can be rewritten as follows:

$$\begin{aligned} aLFI &= \frac{1}{n\Delta\theta^2} \sum_{i=1}^N \frac{(df^T \xi_i)^2}{\lambda_i} \\ &= \frac{1}{n\Delta\theta^2} \frac{|df|^2}{\lambda} \sum_{i=1}^N \frac{(df^T \xi_i)^2}{\lambda_i}, \end{aligned} \quad (4)$$

where  $\lambda$  is the mean variance, and  $\lambda_i = \lambda \times \hat{\lambda}_i$ .  $\hat{df} = \frac{df}{|df|}$  is the unit vector with length of 1 and direction as the same as the signal vector  $df$ . According to equation (4), we disentangled the potential mechanisms of improved LFI into four subparts: signal enhancement, reflected by the modulus length  $|df|$ ; manifold shrinkage, reflected by the mean variance of  $\lambda$ ; signal rotation, reflected by the direction of the signal vector  $df$ ; and manifold warping, reflected by the relative angle of both  $\xi_i$  and  $\hat{\lambda}_i$ . We applied a stepwise approach to assess their respective contributions by sequentially allowing only one mechanism to occur and calculating the resulting changes in aLFI. Specifically, we first calculated aLFI at pre-test as

$$aLFI_{pre} = \frac{1}{n\Delta\theta^2} \frac{|df_{pre}|^2}{\lambda_{pre}} \sum_{i=1}^N \frac{\left(\frac{df_{pre}^T}{|df_{pre}|} \xi_i^{pre}\right)^2}{\hat{\lambda}_i^{pre}}. \quad (5)$$

Considering only the effect of signal enhancement, we can calculate its effect as

$$aLFI_{se} = \frac{1}{n\Delta\theta^2} \frac{|df_{post}|^2}{\lambda_{pre}} \sum_{i=1}^N \frac{\left(\frac{df_{pre}^T}{|df_{pre}|} \xi_i^{pre}\right)^2}{\hat{\lambda}_i^{pre}}. \quad (6)$$

Note that the only difference here is that the  $|df_{pre}|^2$  in equation (5) is replaced by the  $|df_{post}|^2$  in equation (6). The difference between  $aLFI_{se}$  and  $aLFI_{pre}$  is considered as the information gain introduced by the signal enhancement mechanism (that is, the brown bars in Fig. 4b). Following this idea, we can calculate the stepwise aLFI by one-by-one considering the effects of manifold shrinkage  $aLFI_{ms}$ , signal rotation  $aLFI_{sr}$  and manifold warping ( $aLFI_{mw}$  or  $aLFI_{post}$ ) as

$$aLFI_{ms} = \frac{1}{n\Delta\theta^2} \frac{|df_{post}|^2}{\lambda_{post}} \sum_{i=1}^N \frac{\left(\frac{df_{pre}^T}{|df_{pre}|} \xi_i^{pre}\right)^2}{\hat{\lambda}_i^{pre}} \quad (7)$$

$$\text{aLFI}_{\text{sr}} = \frac{1}{n\Delta\theta^2} \frac{|\mathbf{df}_{\text{post}}|^2}{\lambda_{\text{post}}} \sum_{i=1}^N \frac{\left( \frac{\mathbf{df}_{\text{post}}^T}{|\mathbf{df}_{\text{post}}|} \xi_i^{\text{pre}} \right)^2}{\lambda_i^{\text{pre}}} \quad (8)$$

$$\text{aLFI}_{\text{mw}} = \text{aLFI}_{\text{post}} = \frac{1}{n\Delta\theta^2} \frac{|\mathbf{df}_{\text{post}}|^2}{\lambda_{\text{post}}} \sum_{i=1}^N \frac{\left( \frac{\mathbf{df}_{\text{post}}^T}{|\mathbf{df}_{\text{post}}|} \xi_i^{\text{post}} \right)^2}{\lambda_i^{\text{post}}}. \quad (9)$$

The information gain in Fig. 4b indicates the difference between  $\text{aLFI}_{\text{se}}$ ,  $\text{aLFI}_{\text{ms}}$ ,  $\text{aLFI}_{\text{sr}}$  and  $\text{aLFI}_{\text{mw}}$  (that is,  $\text{aLFI}_{\text{post}}$ ) as compared with the pre-test baseline  $\text{aLFI}_{\text{pre}}$ . They are shown as brown, blue, green and magenta bars in Fig. 4b, respectively.

### DCNN modelling of motion VPL

**Stimuli.** The experiment used random dot motion (RDM) stimuli, which consist of a cloud of independent moving dots with some degree of coherence in a given moving direction<sup>34</sup>. The network was trained to discriminate whether the moving direction of a target RDM stimulus was 4° clockwise or counterclockwise relative to its corresponding reference RDM stimulus. To meet the network's specifications, the motion stimuli were 16-frame videos ( $112 \times 112$  pixels per frame). Within each frame, ~100 dots were displayed, with each dot represented by a cross of 3 pixels in both height and width. We set eight coherence levels (8.84%, 12.5%, 17.7%, 25%, 35.3%, 50%, 70.7% and 100%) and four reference directions (45°, 135°, 225° and 315°). The motion speed was 7.5 pixels per frame. All non-coherently moving dots appeared randomly in the image. The display of each frame was limited to a centred circle with a diameter of 112 pixels, with the surrounding areas displayed in black.

**Neural network architecture and training.** Our DCNN is a three-dimensional (3D) convolutional neural network inherited from the C3D network for action recognition<sup>36</sup>. The original C3D consists of ten convolutional layers and three fully connected layers. The main difference between C3D and AlexNet is that C3D uses 3D convolutional kernels to process spatiotemporal information. We kept the first six convolutional layers from the pretrained C3D and replaced the three fully connected layers with a fully connected layer that outputs a single scalar. The number of layers was chosen to (1) keep roughly similar number of parameters to the orientation DCNN and (2) to roughly match the number of regions of interest (ROIs) in the human neuroimaging experiment. Similar to the orientation DCNN, the motion DCNN was also configured in a Siamese fashion to perform the two-alternative forced-choice task based on the sigmoid function.

Similar to the orientation DCNN, the entire training procedure consisted of two phases: the pretraining phase and the VPL phase. During the pretraining phase, the network was trained on full-coherence noiseless RDM pairs, whereas during the VPL phase, the network was trained on stimulus pairs across all coherence levels (eight levels). The network was trained for 1,000 epochs in the pretraining phase and 2,000 epochs in the training phase using stochastic gradient descent with a learning rate of  $10^{-7}$ , momentum of 0.9 and weight decay of 0.0005. The parameters were updated to minimize the cross-entropy loss between the network outputs and the true stimulus labels. The initial parameters in the fully connected layer were normally randomized, whereas those in the convolutional layers were taken directly from a pretrained C3D available at [https://download.openmmlab.com/mmaction/recognition/c3d/c3d\\_sports1m\\_16x1x1\\_45e\\_ucf101\\_rgb/c3d\\_sports1m\\_16x1x1\\_45e\\_ucf101\\_rgb\\_20201021-26655025.pth](https://download.openmmlab.com/mmaction/recognition/c3d/c3d_sports1m_16x1x1_45e_ucf101_rgb/c3d_sports1m_16x1x1_45e_ucf101_rgb_20201021-26655025.pth). The entire procedure was repeated ten times for each reference direction to control for randomness. All model and training procedures were implemented using Python 3.10.9 conda environment, including pytorch 1.13.1, scikit-learn 1.2.0. Details of full conda environment are provided via GitHub at [https://github.com/Yu-AngCheng/neural\\_geometry\\_VPL](https://github.com/Yu-AngCheng/neural_geometry_VPL).

**Behavioural and neural analyses.** The behavioural performance of the network was also evaluated by its classification confidence (equation 1) at all coherence levels before and after the visual training phase. In addition, the firing rates of artificial neurons were measured on each trial as the output of the ReLU layers, averaged over all locations and timepoints. All measurements were taken over 1,000 simulated trials. To ensure that units were truly driven by the stimuli, only units with a mean firing rate greater than 0.001 before and after training were included in subsequent analyses.

To perform decoding analyses, we trained a linear classifier on the firing rates of the artificial neurons to discriminate between the target and the reference stimuli. To assess the performance of the classifier, we split all trials half–half as training and test datasets, and used the average performance of the test-set. For comparison with the electrophysiological data, we calculated the Fano factor of each unit as the ratio of the variance of the firing rate to its mean, and the noise correlations as the correlation between the firing rates of units when viewing the same RDM stimulus. In addition, to measure how much information was contained in a layer per unit, we calculated the aLFI (see above).

We further validated the computational mechanisms in the motion direction discrimination task. To this end, the firing rates of the same group of units under the reference and the target stimuli were also considered as two distributions in a high-dimensional neural space. In the high-dimensional neural space, we defined signal vector, signal separation, variance, correlation, signal rotation angle, principal component (PC) strength and PC rotation angle as above.

Again, we computed linear Fisher information using a stepwise approach. For all models, we sequentially added signal enhancement, manifold shrinkage, signal rotation and manifold warping to the calculation of linear Fisher information and examined how the information within units varied with all four mechanisms. Figure 5i shows the results of the stepwise analysis in layer 6. Supplementary Fig. 3 shows the results in all six layers of the motion DCNN.

### Human fMRI experiment

The human fMRI experiment data have been published in ref. 37 for different research questions. The core analyses in this study beyond preprocessing and ROI definitions are specifically designed in this study. We provide relevant methods as follows and more detailed methods in Supplementary Note 4 to avoid cross-referencing.

**Subjects and experimental procedures.** A total of 22 human subjects (10 males and 12 females, ages 17–25 years) participated in the experiment. All participants had normal or correct-to-normal vision. All participants provided written informed consent, and the study obtained approval from the local ethics committee at Peking University (protocol number 2012-03-09). This study was not preregistered. All subjects were compensated 20 yuan and 100 yuan for an hour of behavioural and fMRI experiments, respectively. All participants were blinded to the study's objectives.

All subjects were trained on a direction discrimination task (Fig. 6a; see Supplementary Note 4 for apparatus and stimulus details). The whole experiment consisted of three phases: pre-test (2 days), training (10 days) and post-test (2 days). On day 1 at pre-test and day 2 at post-test, subjects were tested on direction discrimination around 45° and 135° (angular difference 4°, 120 trials for each direction) to assess their behavioural performance before and after training. Subjects were trained on the fine-direction discrimination task for 10 days. Half of the subjects were trained at 45° and the other half at 135° (see training details in Supplementary Note 4). The assignments were randomized across subjects. Training-induced behavioural improvements have been reported in our previous work<sup>37</sup>. All visual stimuli were generated and presented via Psychtoolbox 3.0 in MATLAB2013A.

To assess the neural changes induced by visual training, two identical fMRI sessions were performed on day 1 at pre-test and day 2 at post-test, respectively. In each fMRI session, subjects completed four runs of the motion direction discrimination task. Each run contained 30 trials for 45° and 135° (that is, a total of 120 trials for each direction). Each run also contained 15 fixation trials, and the trial order was randomized.

**MRI data acquisition.** All MRI data were acquired using a 12-channel phase array coil on a Siemens Trio 3T scanner at Peking University. The T1-weighted anatomical data with a resolution of  $1 \times 1 \times 1 \text{ mm}^3$  were collected for each subject. Echo-planar imaging (EPI) functional data were collected for the motion direction discrimination task, retinotopic mapping and motion localizer experiments. EPI data were acquired using gradient echo-pulse sequences from 33 axial slices, covering the whole brain. The standard EPI sequence used for data acquisition was as follows: a repetition time of 2,000 ms, an echo time of 30 ms, a flip angle of 90° and a resolution of  $3 \times 3 \times 3 \text{ mm}^3$ . The slice order was interleaved ascending.

In addition to the four runs of the motion direction discrimination task, we also collected one or two retinotopic mapping runs<sup>52,55</sup> and a motion localizer run<sup>55</sup> to define ROIs.

**MRI data analyses.** In Brain Voyager QX (version 2.8.0), the anatomical data were transformed into the Talairach coordinate space. For all functional data, the first four volumes of each functional run were discarded to allow the longitudinal magnetization to reach a steady state. The functional data underwent several standard preprocessing procedures, including slice timing correction, head motion correction, spatial smoothing, temporal high-pass filtering (generalized linear model (GLM) with Fourier basis set at two cycles) and linear trend removal. Brain Voyager QX (version 2.8.0) was also used to preprocess the data of the retinotopic mapping experiment and the motion localizer experiment. We used the standard phase-encoding method to define the retinotopic visual areas V1, V2, V3 and V3A (refs. 56,57). A GLM was then applied to the motion localizer data to define the motion-selective voxels (hMT+ and motion-selective voxels in IPS).

The functional data of the motion direction discrimination task were preprocessed using SPM12 ([www.fil.ion.ucl.ac.uk/spm](http://www.fil.ion.ucl.ac.uk/spm)). The data were aligned to the first volume of the first run of the first session, corrected for acquisition delay and then normalized to the Montreal Neurological Institute (MNI) coordinate space using an EPI template. We used the GLMdenoise package (version 1.4, <http://www.kendrickkay.net/GLMdenoise/>) developed in ref. 58 without evoking multirun denoise procedures to estimate the single-trial activity of voxels.

**Voxel population response analyses.** We adapted the analysis previously used for artificial neurons in neural networks to the single-trial fMRI response estimates. To improve SNR, we selected the 60 most responsive voxels in each ROI at pre-test. We first investigated which ROI was involved in motion VPL by measuring the discriminability between two different motion conditions (trained direction, for example, 45° versus untrained direction, for example, 135°) before and after training. We trained a linear classifier on the fMRI data to discriminate between the two motion conditions. To assess the performance of the classifier, we performed a leave-one-trial-out cross-validation, and the average performance on the leave-out test trial was used as the discriminability measure. We also computed the average linear Fisher information (see equations above) between the 45° versus 135° conditions to quantify stimulus discriminability. We found that motion direction discrimination training significantly improved stimulus discriminability in V3A and hMT+. Therefore, we included only V3A and hMT+ voxels in the subsequently analyses.

Similar to the analyses in the DCNNs, we defined the signal vector, the signal separation, the variance, the intervoxel correlations, the signal rotation angle, the PC strength and the PC rotation angle in the multivoxel high-dimensional space using the same method defined above (Fig. 6). In addition, we applied the same stepwise analysis approach of calculating aLFI to the fMRI data (Fig. 6k).

### Monkey multiunit recording experiment

Part of the monkey psychophysical and neurophysiological data have been published in refs. 14,59. These previous studies showed qualitatively similar results of the learning-induced reduction in Fisher information, Fano factor and noise correlations via different analysis methods. Other results and analyses on the characteristics of population responses in this study (that is, Figs. 7 and 8), especially the validation of signal rotation and manifold warping mechanisms, as well as the stepwise information analyses, are key contribution of our study. We provide relevant methods as follows and more detailed methods in Supplementary Note 5 to avoid cross-referencing.

**Ethics statement and data collection.** The Newcastle University Animal Welfare Ethical Review Board approved all procedures in this study. All experimental procedures were carried out in accordance with the European Communities Council Directive RL 2010/63/EC, the US National Institutes of Health Guidelines for the Care and Use of Animals for Experimental Procedures and the UK Animals Scientific Procedures Act. This study included two male monkey monkeys (5 and 14 years of age). This study was not preregistered. ARRIVE guidelines were used to report the research.

**Experimental preparation.** The surgical procedure is described in ref. 60 and Supplementary Note 5. The headpost and electrode implementations are also described in Supplementary Note 5. In brief, in monkey 1, two  $4 \times 5$  grids of microelectrodes were implanted in area V4; in monkey 2, one  $5 \times 5$  grid was implanted in V4. These chronically implanted electrodes allowed us to record population activity in area V4 over the course of visual training. Importantly, we were able to record stably from a few small multiunit clusters. The stability of the recording is shown in ref. 14. Stable recording of multichannel neuronal activity allows analyses of changes in population responses induced by training.

**Behavioural task and monkey training.** All monkey training and data collections were conducted by CORTEX software (last updated in 2013, <http://dally.nimh.nih.gov/index.html>). The monkeys were trained in a contrast discrimination task in which subjects were asked to decide whether the contrast of a test stimulus was higher or lower as compared with that of a reference stimulus by making a saccade to one of two distinct locations (Fig. 7b). On each trial, the subject first kept fixation on the centre of the screen for 512 ms. After 539 ms of fixation, a vertically oriented reference Gabor stimulus with 30% contrast was presented, centred at the V4 receptive field coordinates. The outer diameter of the Gabor stimulus was truncated at 16° for monkey 1 and 14° for monkey 2. After the Gabor stimulus, monkey 2 experienced an interstimulus interval of 512 ms. By contrast, monkey 1 experienced a randomly chosen interstimulus interval, ranging from 512 to 1,024 ms. During the interstimulus interval, only the fixation dot was presented. A test stimulus was then presented for 512 ms. This test stimulus was identical in size and orientation to the reference stimulus but differed in contrast, with the contrast level chosen pseudorandomly. The test stimulus was followed by another blank period of 512 ms during which only the fixation dot was visible. After the fixation dot, two target squares, one black and one white with a size of 0.5° in size, appeared to the left and right of the location where the reference and test stimuli were previously presented. The monkeys were cued to make a decision once the fixation dot disappeared. The monkeys were required to make a saccade to the white square within a 2° × 2° window if the test stimulus

had a higher contrast than the reference stimulus. Conversely, they were expected to make a saccade to the black square if the test stimulus had a lower contrast than the reference stimuli. A correct saccade was rewarded with a fluid reward, while an incorrect saccade led to no reward and a 0.2 s timeout period.

The two monkeys were first trained on an easy version (target contrast 5% or 90%) of the contrast discrimination task. After they were fully familiar with the easy task, the target contrast increased from 2 to 8, 12 and 14 levels. The data correspond to the 14 levels of target contrast (10%, 15%, 20%, 25%, 27%, 28%, 29%, 31%, 32%, 33%, 35%, 40%, 50% or 60%; Supplementary Note 5). We focus only on target contrast levels (27%, 28%, 29%, 31%, 32% and 33%) near the reference contrast (that is, 30%) according to the definition of linear Fisher information.

**Dataset and preprocessing.** We used chronically implanted Utah arrays to record spiking activity. We refer to small multiunit neuronal clusters recorded from a given electrode as channels. Twenty-nine and 20 channels were recorded in monkey 1 and monkey 2, respectively. These channels exhibited good responses ( $\text{SNR} > 1$ ) on over 80% of the recording sessions (see SNR computation in Supplementary Note 5). Baseline activity matching was performed between sessions for multiunit activity data to obtain comparable activity levels across sessions.

**Behavioural and neural analyses.** We noticed that the relationship between neural activity and discriminability can change drastically during the stimulus presentation period, and through training, the improvement in discriminability can also vary over the course of the training period. We chose the first four and the last four training sessions as the early and the late phase of training. This choice ensures an overall sufficient and comparable number of trials at both pre- and post-test for further analyses.

To determine the time window, we systematically varied the time window and trained a linear classifier to discriminate between the reference and target stimuli, and obtained its performance through tenfold cross-validation. We chose the time window with the largest change in decoding accuracy between the reference stimulus (30% contrast) and the target stimuli (29% or 31% contrast). For monkey 1, the chosen time window was 30–130 ms after stimulus onset. For monkey 2, the time window was 130–230 ms after stimulus onset. Note that this choice aims to maximize training effects on population representations (similar to the decoding analyses for first identifying V3A and hMT<sup>+</sup> as the ROIs where learning effects are most pronounced in the human fMRI study) but does not guarantee the underlying mechanisms such as signal separation enhancement and manifold shrinkage. Also, varying the time window did not qualitatively change our results. We used a simple multivariate Poisson log-normal model (Supplementary Note 5, see also refs. 61–64) to estimate the trial-by-trial variability of population firing rates. We further use the estimated firing rates and covariance to compute all neural metrics mentioned above. We report all results in Figs. 7 and 8 for visual comparison with the DCNN and fMRI results above.

## Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

All data to reproduce the figures in the Article and its Supplementary Information are available via GitHub at [https://github.com/Yu-AngCheng/neural\\_geometry\\_VPL](https://github.com/Yu-AngCheng/neural_geometry_VPL). The raw human fMRI and monkey physiological data used in this study were all published previously<sup>14,37</sup>. Requests for other datasets should be directed to the original authors who collected the data.

## Code availability

The code for training neural networks, stimulus generation and neural geometry analysis is publicly available via GitHub at [https://github.com/Yu-AngCheng/neural\\_geometry\\_VPL](https://github.com/Yu-AngCheng/neural_geometry_VPL).

## References

1. Lu, Z. L., Yu, C., Watanabe, T., Sagi, D. & Levi, D. Perceptual learning: functions, mechanisms, and applications. *Vis. Res.* **50**, 365–367 (2010).
2. Watanabe, T. & Sasaki, Y. Perceptual learning: toward a comprehensive theory. *Annu. Rev. Psychol.* **66**, 197–221 (2015).
3. Chen, N. et al. Sharpened cortical tuning and enhanced cortico-cortical communication contribute to the long-term neural mechanisms of visual motion perceptual learning. *NeuroImage* **115**, 17–29 (2015).
4. Shibata, K. et al. Decoding reveals plasticity in V3A as a result of motion perceptual learning. *PLoS ONE* **7**, e44003 (2012).
5. Jia, K. et al. Recurrent processing drives perceptual plasticity. *Curr Biol.* **30**, 4177–4187 (2020).
6. Lu, Z. L. & Dosher, B. A. Characterizing observers using external noise and observer models: assessing internal representations with external noise. *Psychol Rev.* **115**, 44–82 (2008).
7. Dosher, B. A. & Lu, Z. L. Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proc. Natl Acad. Sci. USA* **95**, 13988–13993 (1998).
8. Dosher, B. & Lu, Z. L. Visual perceptual learning and models. *Annu. Rev. Vis. Sci.* **3**, 343–363 (2017).
9. Schoups, A., Vogels, R., Qian, N. & Orban, G. Practising orientation identification improves orientation coding in V1 neurons. *Nature* **412**, 549–553 (2001).
10. Yang, T. & Maunsell, J. H. The effect of perceptual learning on neuronal responses in monkey visual area V4. *J. Neurosci.* **24**, 1617–1626 (2004).
11. Ni, A. M., Ruff, D. A., Alberts, J. J., Symmonds, J. & Cohen, M. R. Learning and attention reveal a general relationship between population activity and behavior. *Science* **359**, 463–465 (2018).
12. Gu, Y. et al. Perceptual learning reduces interneuronal correlations in macaque visual cortex. *Neuron* **71**, 750–761 (2011).
13. Yan, Y. et al. Perceptual training continuously refines neuronal population codes in primary visual cortex. *Nat. Neurosci.* **17**, 1380–1387 (2014).
14. Sanaye, M. et al. Perceptual learning of fine contrast discrimination changes neuronal tuning and population coding in macaque V4. *Nat. Commun.* **9**, 4238 (2018).
15. Kohn, A., Coen-Cagli, R., Kanitscheider, I. & Pouget, A. Correlations and neuronal population information. *Annu. Rev. Neurosci.* **39**, 237–256 (2016).
16. Moreno-Bote, R. et al. Information-limiting correlations. *Nat. Neurosci.* **17**, 1410–1417 (2014).
17. Zhang, R. Y., Wei, X. X. & Kay, K. Understanding multivariate brain activity: evaluating the effect of voxelwise noise correlations on population codes in functional magnetic resonance imaging. *PLoS Comput. Biol.* **16**, e1008153 (2020).
18. Cheng, Z. J., Yang, L., Zhang, W. H. & Zhang, R. Y. Representational geometries reveal differential effects of response correlations on population codes in neurophysiology and functional magnetic resonance imaging. *J. Neurosci.* **43**, 4498–4512 (2023).
19. Wenliang, L. K. & Seitz, A. R. Deep neural networks for modeling visual perceptual learning. *J. Neurosci.* **38**, 6028–6044 (2018).
20. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **25**, 1097–1105 (2012).
21. Dosher, B. A. & Lu, Z. L. Mechanisms of perceptual learning. *Vis. Res.* **39**, 3197–3221 (1999).

22. Chen, N., Cai, P., Zhou, T., Thompson, B. & Fang, F. Perceptual learning modifies the functional specializations of visual cortical areas. *Proc. Natl Acad. Sci. USA* **113**, 5724–5729 (2016).
23. Raiguel, S., Vogels, R., Mysore, S. G. & Orban, G. A. Learning to see the difference specifically alters the most informative V4 neurons. *J. Neurosci.* **26**, 6589–6602 (2006).
24. Ghose, G. M., Yang, T. & Maunsell, J. H. Physiological correlates of perceptual learning in monkey V1 and V2. *J. Neurophysiol.* **87**, 1867–1888 (2002).
25. Jehee, J. F., Ling, S., Swisher, J. D., van Bergen, R. S. & Tong, F. Perceptual learning selectively refines orientation representations in early visual cortex. *J. Neurosci.* **32**, 16747–16753a (2012).
26. Jeanne, J. M., Sharpee, T. O. & Gentner, T. Q. Associative learning enhances population coding by inverting interneuronal correlation patterns. *Neuron* **78**, 352–363 (2013).
27. Ruff, D. A. & Cohen, M. R. Attention can either increase or decrease spike count correlations in visual cortex. *Nat. Neurosci.* **17**, 1591–1597 (2014).
28. Hennequin, G., Ahmadian, Y., Rubin, D. B., Lengyel, M. & Miller, K. D. The dynamical regime of sensory cortex: stable dynamics around a single stimulus-tuned attractor account for patterns of noise variability. *Neuron* **98**, 846–860 (2018).
29. Ecker, A. S. et al. Decorrelated neuronal firing in cortical microcircuits. *Science* **327**, 584–587 (2010).
30. Gutnisky, D. A. & Dragoi, V. Adaptive coding of visual information in neural populations. *Nature* **452**, 220–224 (2008).
31. Huang, X. & Lisberger, S. G. Noise correlations in cortical area MT and their potential impact on trial-by-trial variation in the direction and speed of smooth-pursuit eye movements. *J. Neurophysiol.* **101**, 3012–3030 (2009).
32. van Bergen, R. S. & Jehee, J. F. M. Modeling correlated noise is necessary to decode uncertainty. *NeuroImage* **180**, 78–87 (2018).
33. Liu, Z. Perceptual learning in motion discrimination that generalizes across motion directions. *Proc. Natl Acad. Sci. USA* **96**, 14085–14087 (1999).
34. Ball, K. & Sekuler, R. Direction-specific improvement in motion discrimination. *Vis. Res.* **27**, 953–965 (1987).
35. Law, C. T. & Gold, J. I. Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. *Nat. Neurosci.* **11**, 505–513 (2008).
36. Tran, D., Bourdev, L., Fergus, R., Torresani, L. & Paluri, M. Learning spatiotemporal features with 3D convolutional networks. In *Proc. IEEE International Conference on Computer Vision* 4489–4497 (IEEE, 2015).
37. Jia, K. et al. Visual perceptual learning modulates decision network in the human brain: the evidence from psychophysics, modeling, and functional magnetic resonance imaging. *J. Vis.* **18**, 9 (2018).
38. Thomson, H. L. et al. Haemodynamics and left ventricular mass regression: a comparison of the stentless, stented and mechanical aortic valve replacement. *Eur. J. Cardiothorac. Surg.* **13**, 572–575 (1998).
39. Ashby, F. G. Estimating the parameters of multidimensional signal detection theory from simultaneous ratings on separate stimulus components. *Percept. Psychophys.* **44**, 195–204 (1988).
40. Ashby, F. G. & Soto, F. A. in *The Oxford Handbook of Computational and Mathematical Psychology* (eds Busemeyer, J. R. et al.) 13–34 (Oxford Univ. Press, 2015).
41. Kriegeskorte, N., Goebel, R. & Bandettini, P. Information-based functional brain mapping. *Proc. Natl Acad. Sci. USA* **103**, 3863–3868 (2006).
42. Nastase, S. A. et al. Attention selectively reshapes the geometry of distributed semantic representation. *Cereb. Cortex* **27**, 4277–4291 (2017).
43. Kriegeskorte, N. & Wei, X. X. Neural tuning and representational geometry. *Nat. Rev. Neurosci.* **22**, 703–718 (2021).
44. Charest, I., Kievit, R. A., Schmitz, T. W., Deca, D. & Kriegeskorte, N. Unique semantic space in the brain of each beholder predicts perceived similarity. *Proc. Natl Acad. Sci. USA* **111**, 14565–14570 (2014).
45. Khosla, M., Williams, A. H., McDermott, J. & Kanwisher, N. Privileged representational axes in biological and artificial neural networks. Preprint at bioRxiv <https://doi.org/10.1101/2024.06.20.599957> (2024).
46. Bejjanki, V. R., Beck, J. M., Lu, Z. L. & Pouget, A. Perceptual learning as improved probabilistic inference in early sensory areas. *Nat. Neurosci.* **14**, 642–648 (2011).
47. Regan, D. & Beverley, K. I. Postadaptation orientation discrimination. *J. Opt. Soc. Am. A* **2**, 147–155 (1985).
48. Li, W., Piech, V. & Gilbert, C. D. Perceptual learning and top-down influences in primary visual cortex. *Nat. Neurosci.* **7**, 651–657 (2004).
49. Yan, Y., Zhaoping, L. & Li, W. Bottom-up saliency and top-down learning in the primary visual cortex of monkeys. *Proc. Natl Acad. Sci. USA* **115**, 10499–10504 (2018).
50. Frank, S. M. et al. Supervised learning occurs in visual perceptual learning of complex natural images. *Curr. Biol.* **30**, 2995–3000 (2020).
51. Tartaglia, E. M., Bamert, L., Mast, F. W. & Herzog, M. H. Human perceptual learning by mental imagery. *Curr. Biol.* **19**, 2081–2085 (2009).
52. Lu, Z. L. & Dosher, B. A. Current directions in visual perceptual learning. *Nat. Rev. Psychol.* **1**, 654–668 (2022).
53. Lu, Z. L., Chu, W. & Dosher, B. A. Perceptual learning of motion direction discrimination in fovea: separable mechanisms. *Vis. Res.* **46**, 2315–2327 (2006).
54. Britten, K. H., Shadlen, M. N., Newsome, W. T. & Movshon, J. A. The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci.* **12**, 4745–4765 (1992).
55. Huk, A. C., Dougherty, R. F. & Heeger, D. J. Retinotopy and functional subdivision of human areas MT and MST. *J. Neurosci.* **22**, 7195–7205 (2002).
56. Engel, S. A., Glover, G. H. & Wandell, B. A. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb. Cortex* **7**, 181–192 (1997).
57. Sereno, M. I. et al. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* **268**, 889–893 (1995).
58. Kay, K. N., Rokem, A., Winawer, J., Dougherty, R. F. & Wandell, B. A. GLMdenoise: a fast, automated technique for denoising task-based fMRI data. *Front. Neurosci.* **7**, 247 (2013).
59. Chen, X., Sanaye, M. & Thiele, A. Perceptual learning of contrast discrimination in macaca mulatta. *J. Vis.* **13**, 22 (2013).
60. Thiele, A., Delicato, L. S., Roberts, M. J. & Gieselmann, M. A. A novel electrode-pipette design for simultaneous recording of extracellular spikes and iontophoretic drug application in awake behaving monkeys. *J. Neurosci. Methods* **158**, 207–211 (2006).
61. Goris, R. L., Movshon, J. A. & Simoncelli, E. P. Partitioning neuronal variability. *Nat. Neurosci.* **17**, 858–865 (2014).
62. Henaff, O. J. et al. Primary visual cortex straightens natural video trajectories. *Nat. Commun.* **12**, 5982 (2021).
63. Rabinowitz, N. C., Goris, R. L., Cohen, M. & Simoncelli, E. P. Attention stabilizes the shared gain of V4 populations. *eLife* **4**, e08998 (2015).
64. Zhu, R. J. B. & Wei, X. X. Unsupervised approach to decomposing neural tuning variability. *Nat. Commun.* **14**, 2298 (2023).

## Acknowledgements

We thank S. Kuai, D. Tadin and O.-S. Kwon for valuable comments on the manuscripts. This work was supported by the National Science and Technology Innovation 2030 Major Program of China (grant number 2022ZD0204802 to F.F.), National Natural Science Foundation of China (grant numbers 32441102 and 32100901 to R.-Y.Z.; grant numbers T2421004 and 31930053 to F.F.; grant number 3230085 to K.J.), Shanghai Municipal Education Commission (grant number 2024AIZD014 to R.-Y.Z.), the Open Research Fund of the State Key Laboratory of Brain-Machine Intelligence Zhejiang University (grant number BMI2400021 to R.-Y.Z.), the The Fred M Seed Foundation to T.W. The monkey work was supported by the Medical Research Council, UK (grant number G0700976 to A.T.). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

## Author contributions

Y.-A.C. and R.-Y.Z. conceived and designed the study. Y.-A.C. implemented the neural networks. K.J., S.L. and F.F. prepared and provided the preprocessed fMRI data. A.T., X.C. and M.S. recorded, organized and preprocessed the monkey physiological data. Y.-A.C. and R.-Y.Z. performed in-depth analyses on neural networks, human fMRI data and monkey electrophysiological data. Y.-A.C., T.W. and R.-Y.Z. wrote the first draft of the paper. All authors revised the manuscript and provided valuable feedback to the final paper.

## Competing interests

X.C. is a cofounder and shareholder of a neurotechnology start-up, Phosphoenix (the Netherlands). The other authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41562-025-02149-x>.

**Correspondence and requests for materials** should be addressed to Ru-Yuan Zhang.

**Peer review information** *Nature Human Behaviour* thanks Zhong Lin Lu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

<sup>1</sup>Brain Health Institute, National Center for Mental Disorders, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine and School of Psychology, Shanghai, People's Republic of China. <sup>2</sup>Department of Cognitive, Linguistic and Psychological Sciences, Brown University, Providence, RI, USA. <sup>3</sup>Biosciences Institute, Newcastle University, Framlington Place, Newcastle upon Tyne, UK. <sup>4</sup>School of Cognitive Sciences, Institute for Research in Fundamental Sciences, Tehran, Iran. <sup>5</sup>Department of Ophthalmology, University of Pittsburgh, Pittsburgh, PA, USA. <sup>6</sup>Affiliated Mental Health Center and Hangzhou Seventh People's Hospital, Zhejiang University School of Medicine, Hangzhou, People's Republic of China. <sup>7</sup>Liangzhu Laboratory, MOE Frontier Science Center for Brain Science and Brain-machine Integration, State Key Laboratory of Brain-Machine Intelligence, Zhejiang University, Hangzhou, People's Republic of China. <sup>8</sup>NHC and CAMS Key Laboratory of Medical Neurobiology, Zhejiang University, Hangzhou, People's Republic of China. <sup>9</sup>School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, People's Republic of China. <sup>10</sup>IDG/McGovern Institute for Brain Research, Peking University, Beijing, People's Republic of China. <sup>11</sup>Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing, People's Republic of China. <sup>12</sup>Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, People's Republic of China.  e-mail: [ruyuanzhang@gmail.com](mailto:ruyuanzhang@gmail.com)



Supplementary information

<https://doi.org/10.1038/s41562-025-02149-x>

# A neural geometry approach comprehensively explains apparently conflicting models of visual perceptual learning

In the format provided by the authors and unedited

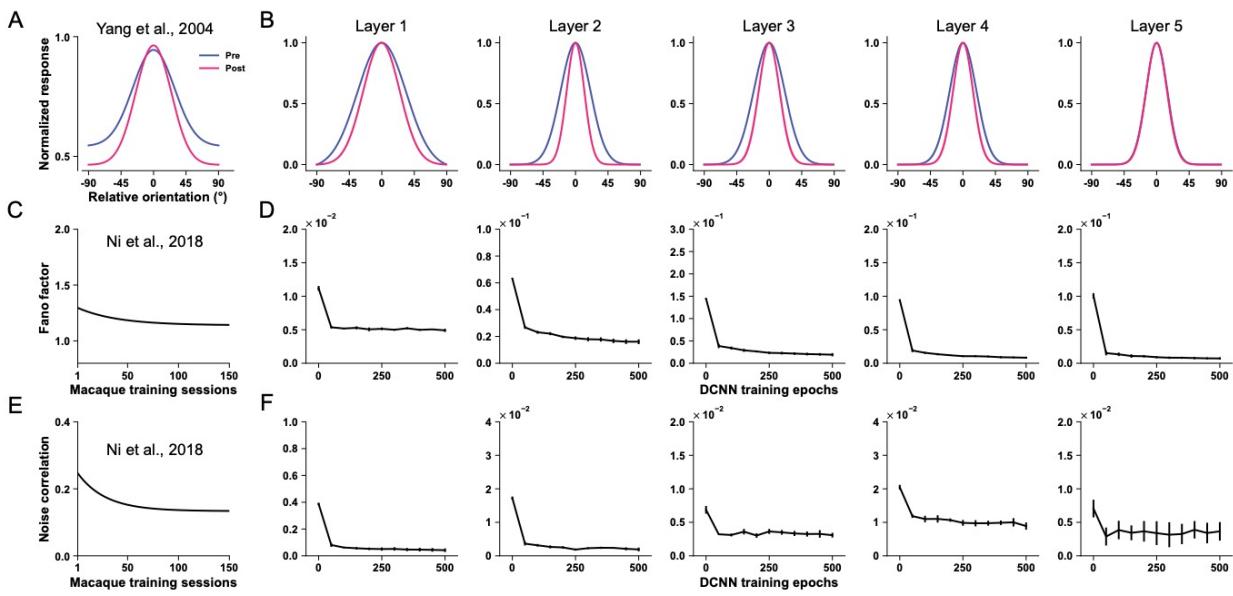
1    **Supplementary Information for**  
2  
3    **A neural geometry approach comprehensively explains apparently conflicting models of**  
4    **visual perceptual learning**

5  
6    **Table of Content**

7    **Supplementary Figure S1:** Neural consequences of orientation VPL in all five layers.  
8    **Supplementary Figure S2:** The orientation DCNN replicates the neuronal response properties  
9    in the primate early visual cortex.  
10   **Supplementary Figure S3:** Neural consequences of motion VPL in all six layers.  
11   **Supplementary Figure S4:** Neural consequences of contrast discrimination learning in two  
12   monkeys.  
13   **Supplementary Figure S5:** The effects of the mechanism order in the stepwise analyses.  
14   **Supplementary Figure S6:** Replications of macaque neurophysiological and human fMRI  
15   experiments on signal enhancement.  
16   **Supplementary Figure S7:** Predicted effects of orientation discrimination learning on  
17   estimation and detection tasks.  
18   **Supplementary Table S1-S7:** Full statistical results of Figure 2F&G, Figure 3F&H, Figure 4B,  
19   and Figure 5C&D in the main text.  
20   **Supplementary Note 1:** Additional explanations on the stepwise analyses  
21   **Supplementary Note 2:** Analytical relation between the high-dimensional population responses  
22   and one-dimensional decision unit responses.  
23   **Supplementary Note 3:** Computer simulations of perceptual consequences predicted by the  
24   neural geometry approach.  
25   **Supplementary Note 4:** Supplementary methods for the human fMRI experiment.  
26   **Supplementary Note 5:** Supplementary methods for the macaque multielectrode recording  
27   experiment.

28  
29

30 **Supplementary Figure S1**

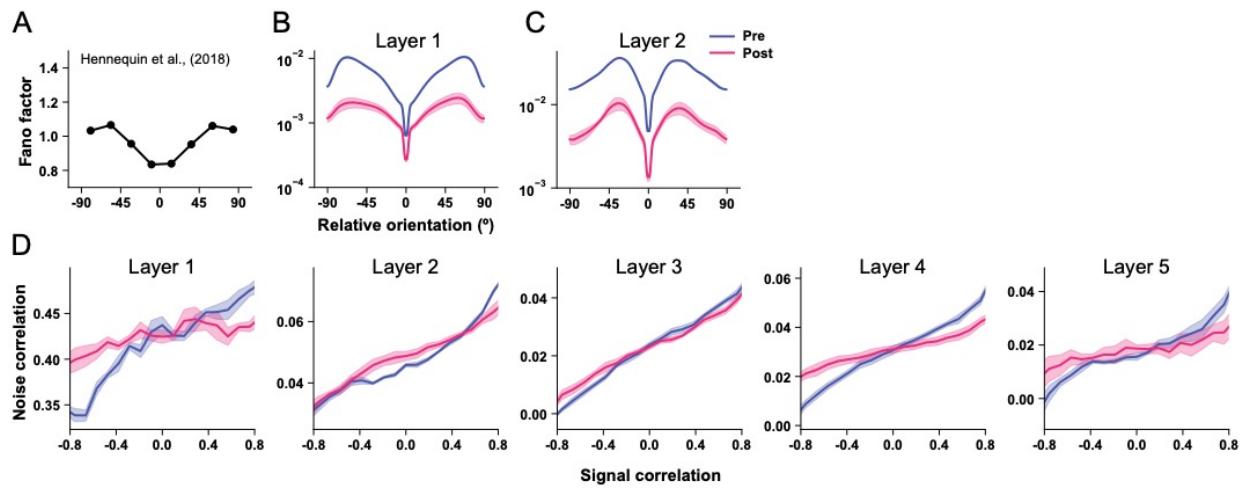


31

32 **Figure S1.** Tuning curve, Fano factor, and noise correlation changes in all five layers of the  
 33 orientation DCNN. Perceptual training on the DCNN reduces the width of the tuning curves in  
 34 layers 1-4. Perceptual training also reduces Fano factor and noise correlations in all five layers.  
 35 Data are presented as mean values  $\pm$  S.E.M., with error bars and error shadings in panels **B-F**  
 36 representing the S.E.M. across four ( $n = 4$ ) reference orientations.

37

38

39 **Supplementary Figure S2**

40

41 **Figure S2.** The DCNN of orientation VPL qualitatively replicates the relationship between  
 42 orientation preferences and Fano factor in macaque V1 (**A**, reproduced from the datapoints from  
 43 ref. 1). Here, the results of layers 1(**B**)&2(**C**) are plotted. The network also reproduces the positive  
 44 relationship between signal correlations and noise correlations in the human [2,3](#) and macaque [4](#)  
 45 early visual systems (**D**). Importantly, interneuron noise correlations are not uniformly reduced by  
 46 training. In contrast, perceptual training reduces the noise correlations between neurons with  
 47 similar tuning preferences (i.e., signal correlation  $> 0$ ) and vice versa between neurons with  
 48 opposite preferences (i.e., signal correlation  $< 0$ ). These results are consistent with several  
 49 theoretical and empirical findings on noise correlation changes induced by learning and attention  
 50 [5-8](#). Data are presented as mean values  $\pm$  S.E.M, with error bars and error shadings in panel **B-D**  
 51 representing the S.E.M. across four ( $n = 4$ ) reference orientations.

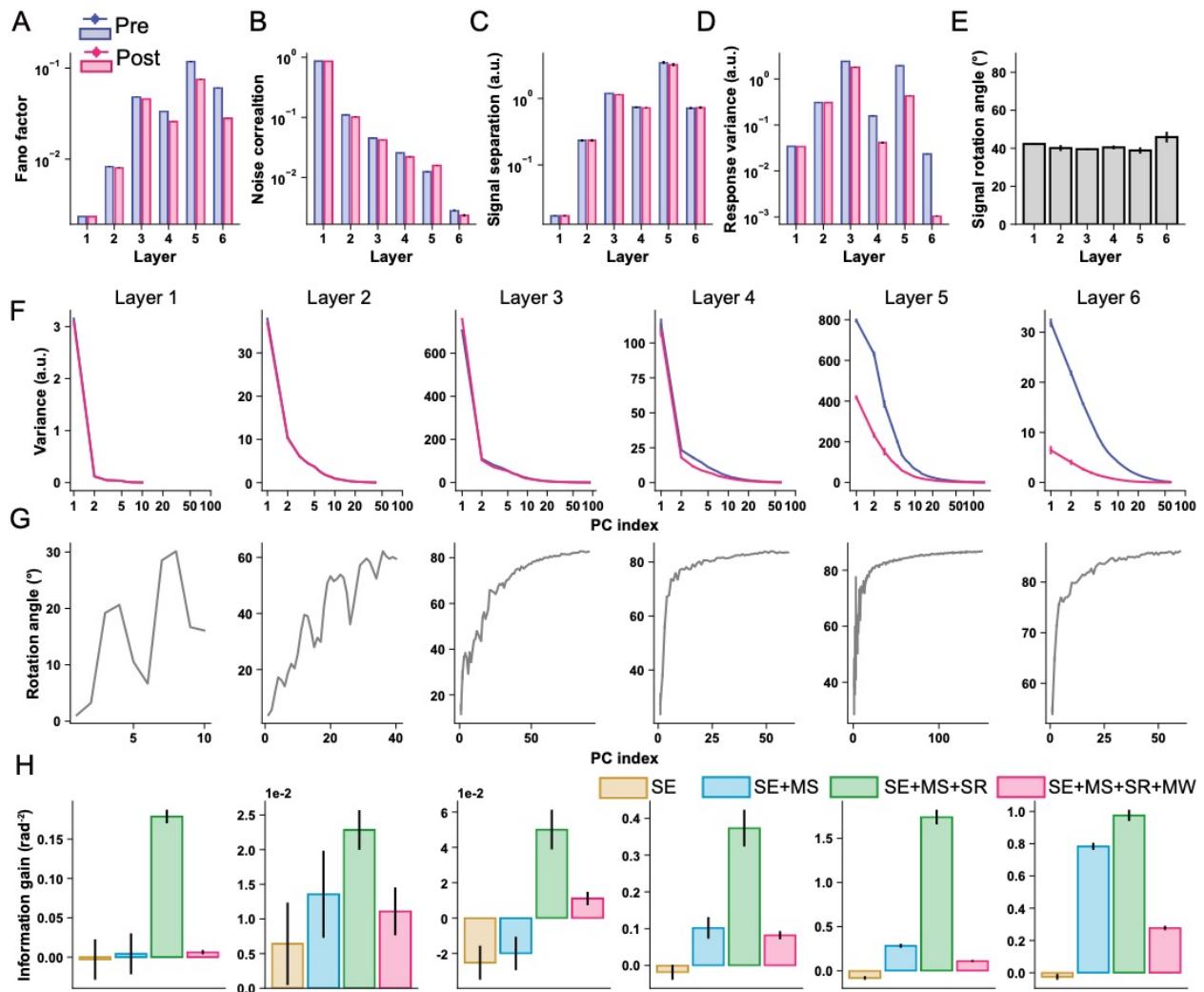
52

53

54

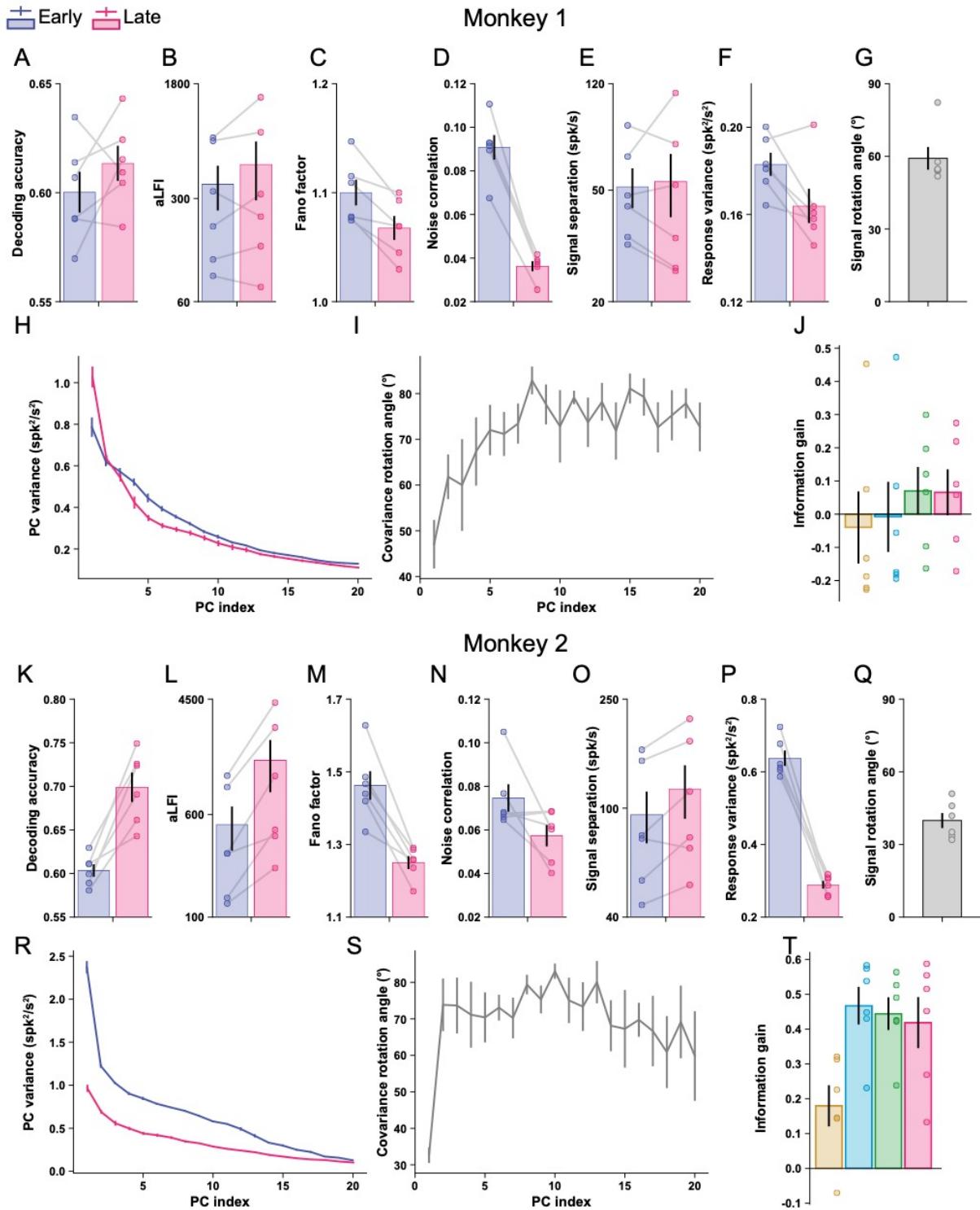
55

56 **Supplementary Figure S3**



57  
58 **Figure S3.** The results of all layers in the motion DCNN. The effects of motion direction  
59 discrimination training on Fano factor (**A**), noise correlations (**B**), signal separation (**C**), response  
60 variance (**D**), and signal rotation angle (**E**) in the DCNN model. The effects of training induced  
61 manifold warping in all six layers are shown in (**F&G**). The results of the stepwise information  
62 analyses are shown in (**H**). Data are presented as mean values  $\pm$  S.E.M., with error bars and error  
63 shadings in panels **A-H** representing the S.E.M. across four ( $n = 4$ ) reference orientations. Note  
64 that some error bars are very small and barely visible.  
65

66 **Supplementary Figure S4**

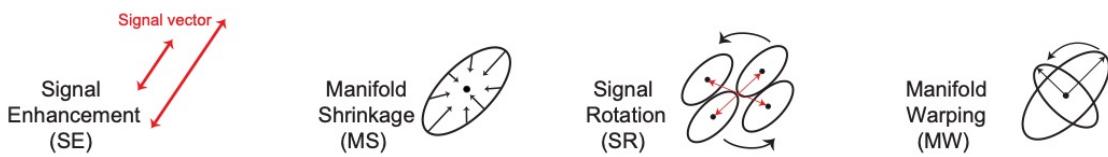


67

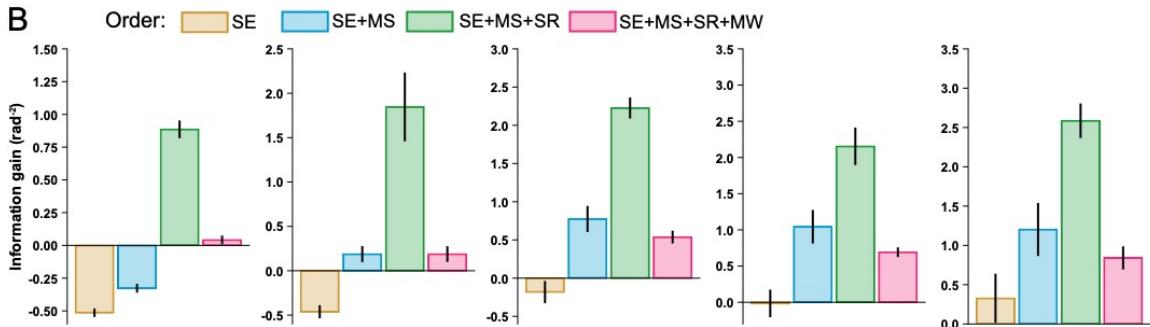
68 **Figure S4.** Population activity analyses for the two monkeys. The organization and conventions  
69 here are the same as Fig. 8 in the main text, except that we split the results for the two monkeys.  
70 Data are presented as mean values  $\pm$  S.E.M., with error bars indicate the S.E.M. across the six  
71 conditions ( $n = 6$ ).  
72

73 Supplementary Figure S5

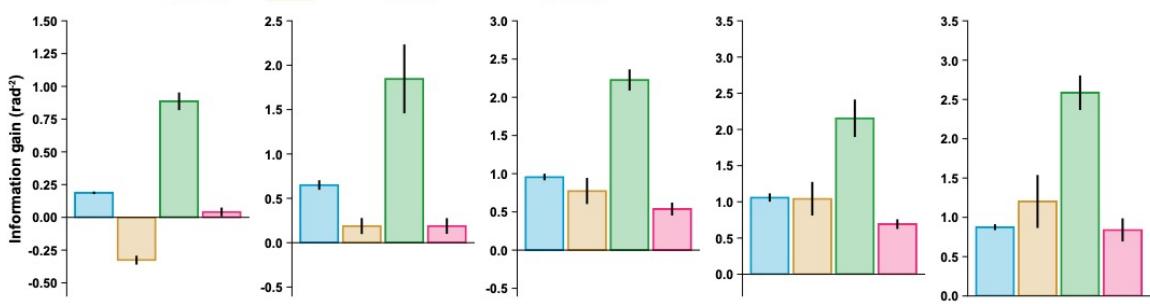
A



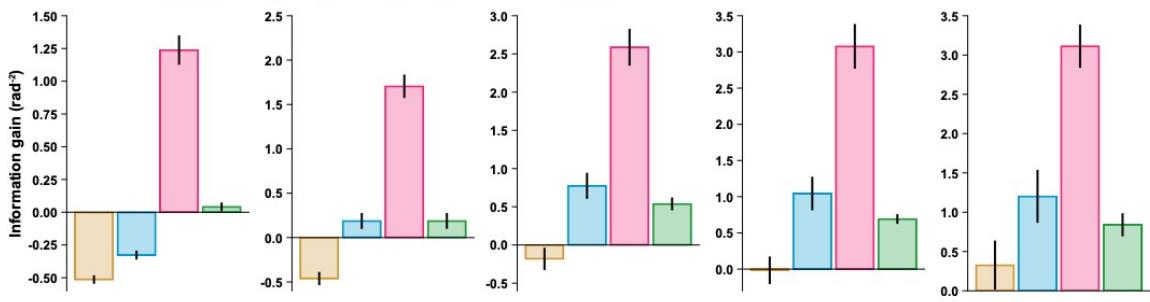
B Layer 1 Layer 2 Layer 3 Layer 4 Layer 5



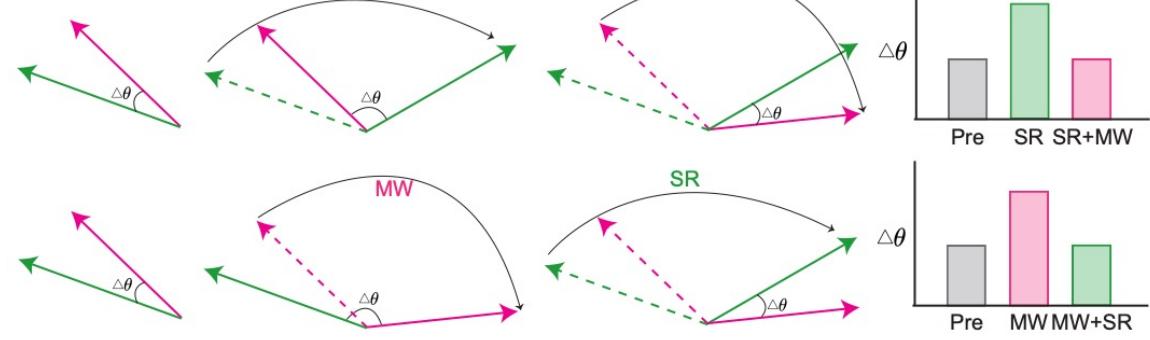
C Order: MS MS+SE MS+SE+SR MS+SE+SR+MW



D Order: SE SE+MS SE+MS+MW SE+MS+MW+SR



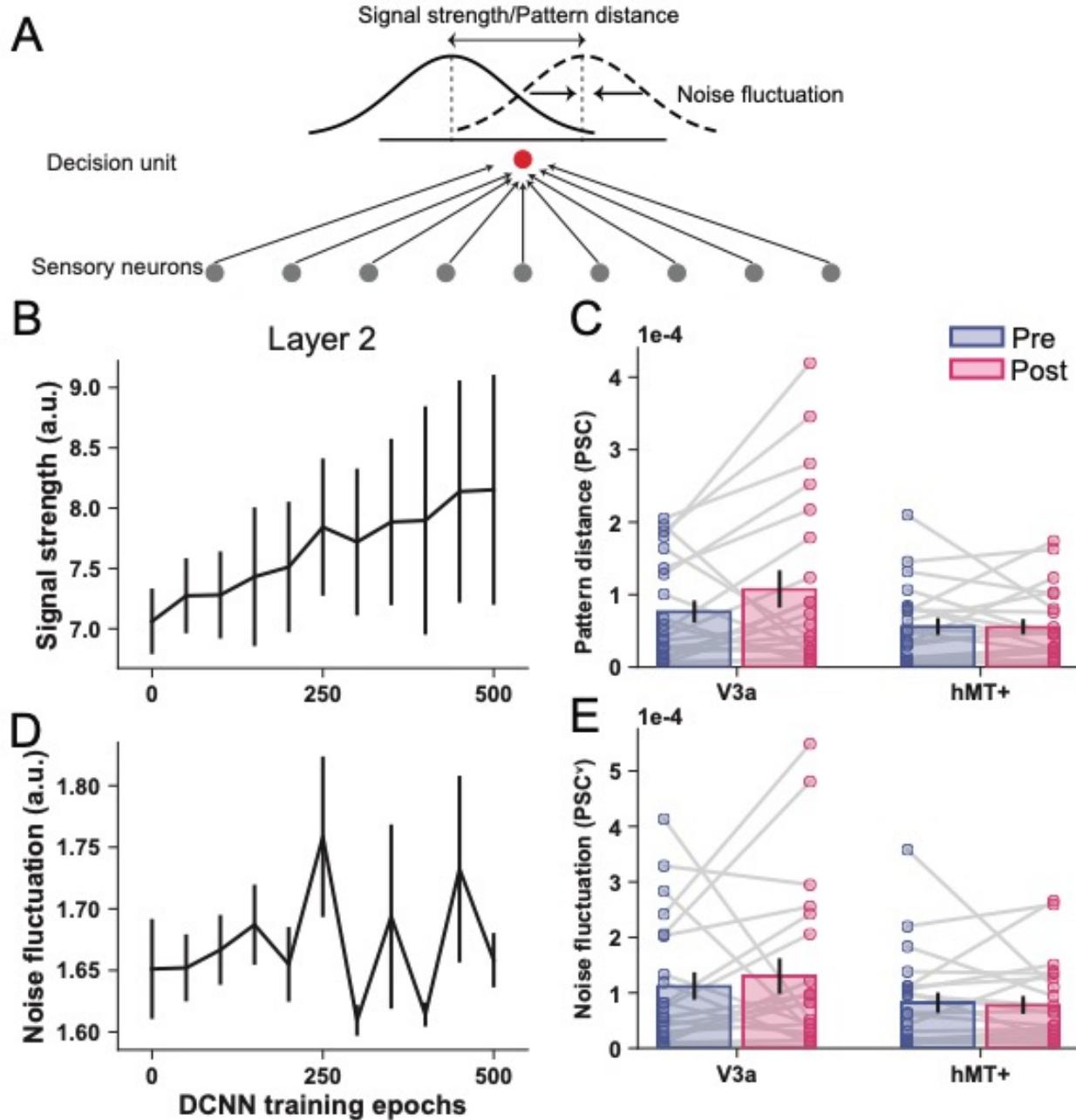
E SR MW SR MW+SR



75 **Figure S5.** The effects of mechanism order on the results of stepwise information analyses. **A&B**.  
76 The cartoon illustration of the four mechanisms (**A**) and the results of the stepwise information  
77 analysis according to the order of SE/SE+MS/SE+MS+SR/SE+MS+SR+MW, the same as in Fig.  
78 4 in the main text. **C**. The results of the stepwise information analysis according to the order of  
79 MS/MS+SE/MS+SE+SR/MS+SE+SR+MW. Here, the mechanisms of SE and MS are  
80 independent. Thus, changing their order does not change the positive effects of MS and the  
81 minimal or even negative effects of SE. More precisely, the information gains of SE and MS are  
82 quantitatively identical in **B** and **C**. For example, in layer 1 of **B**, the height of the yellow bar (~  
83 0.5) indicates the negative effect of SE, which is equal to the decrement from the blue to the yellow  
84 bar in layer 1 of **C**. Similarly, the increment from the yellow to the blue bar in layer 1 of **B** indicates  
85 the positive effect of MS, which is equal to the height of the blue bar in layer 1 of **C**. **D**. Compared  
86 to **B**, we only switch the order of SR and MW in **D** (i.e., the results of SE/SE+MS/SE+MS+MW/  
87 SE+MS+MW+SR). The apparent effects of SR and MW appear to be opposite in **B** and **D**. This is  
88 because what matters here is the projection of the covariance onto the signal vector, or in other  
89 words, the angle between the signal vector and the covariance direction [3·8-10](#). Intuitively, the  
90 larger the angle, the more information the population contains. We explain this effect in cartoon  
91 illustrations in **E**. No matter which mechanism (i.e., SR or MW) is considered first, it increases  
92 the angle between the signal vector and the covariance direction (i.e., increases information), but  
93 this effect is canceled out by further including the other mechanism. Thus, the apparently opposite  
94 effects of SR and MW actually reflect the same underlying mechanism. Data are presented as mean  
95 values  $\pm$  S.E.M., with error bars and error shadings in panels **B-D** representing the S.E.M. across  
96 four ( $n = 4$ ) reference directions.

97

## 98 Supplementary Figure S6



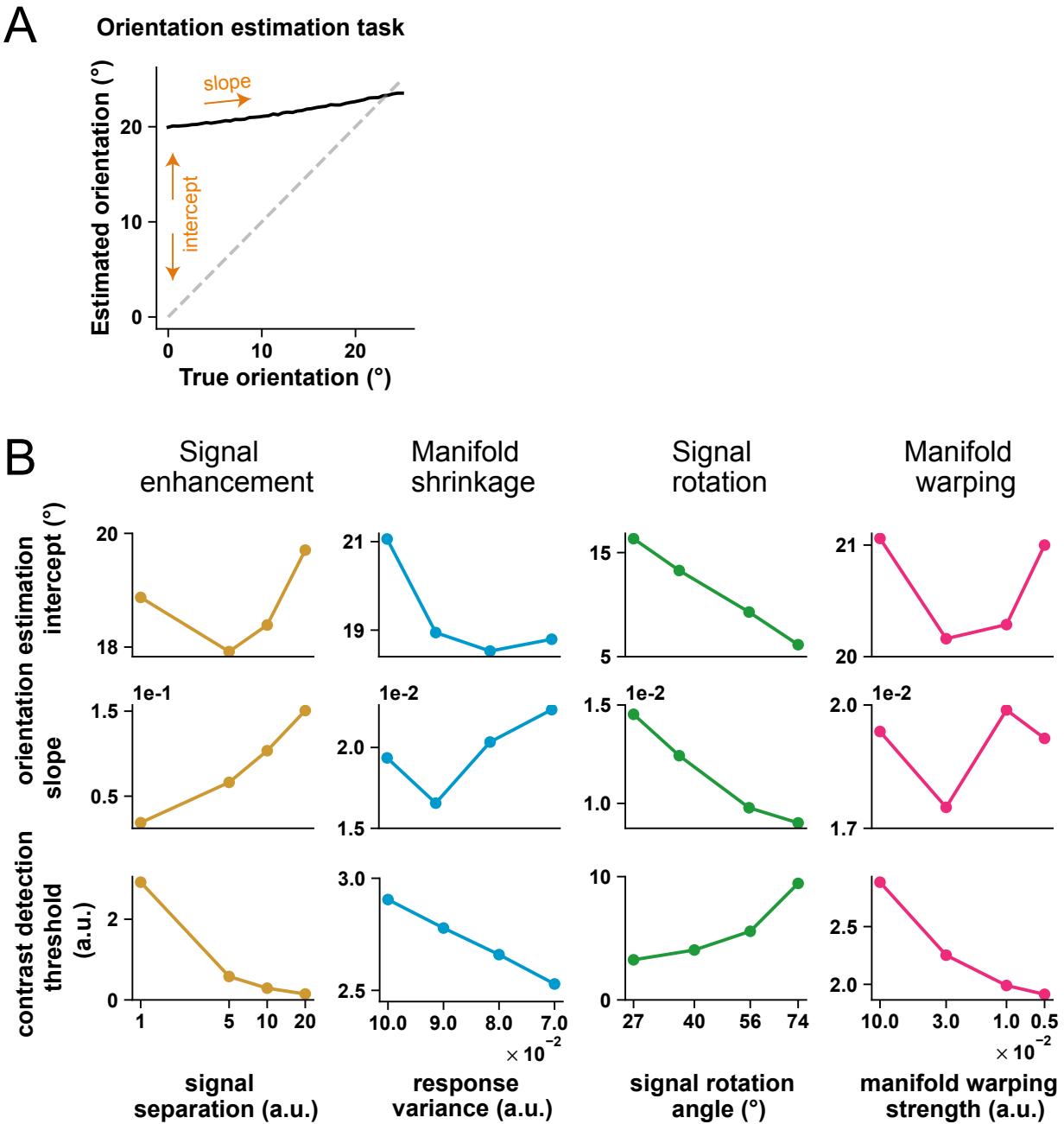
99

100 **Figure S6.** We applied similar analyses in refs. 11-12 to our data. Here, we project the multivariate  
 101 population responses onto an optimal decision unit and derive the two 1-d activity distributions of  
 102 the decision unit for the two stimuli. We replicate the results that, after this linear transformation,  
 103 the effects of VPL appear to be signal enhancement in layer 1 of the orientation DCNN (*A*), and  
 104 in V3A but not in hMT+ in the human brain (*B*). In contrast, noise fluctuation in layer 1 (*C*) and  
 105 in V3A and hMT+ in the human brain (*D*) do not appear to increase. These results are consistent  
 106 with the Fig. 6C in ref. 11 and Fig. 6C in ref. 12. Data are presented as mean values  $\pm$  S.E.M.,  
 107 with error bars and error shadings in panels *B&D* representing the S.E.M. across four ( $n = 4$ )

108 reference orientations and that in panels C&E representing the S.E.M. across twenty-two (n = 22)  
109 subjects.

110

111

112 **Supplementary Figure S7**

113

114

115

116

117

118

119

120

121

122

123

**Figure S7.** Predictions of the neural geometry approach on the effects of VPL of orientation discrimination on perceptual orientation bias reflected in orientation estimation and contrast detection. **A.** The estimated orientation as a function of the true presented orientation (solid line) in an orientation estimation task. In this task, perceptual bias is characterized by the slope and intercept of the estimation function. **B.** The effect of each mechanism in the neural geometry approach (signal enhancement, manifold shrinkage, signal rotation, and manifold warping) on intercept and slope in the orientation estimation task (first and second row, respectively) and contrast detection threshold in the contrast detection task (third row). Note that here our computer simulations focused on learning only on the orientation discrimination task but testing on the orientation estimation and the contrast detection tasks (see details in Supplementary Note 3). The

124 overall trends indicate the magnitude of perceptual consequences (intercept and slope for  
125 perceptual orientation bias and threshold for contrast detection) varies as a function of the strength  
126 of each mechanism as the independent variable, induced by orientation discrimination learning.  
127

128 **Supplementary Table S1.** Full statistical results of the **Figure. 2F** in the main text. All statistical  
 129 results here are obtained by one-side t-tests.

| Layers | $t_3$ | p     | Cohen's d | 95% confidence interval          |
|--------|-------|-------|-----------|----------------------------------|
| 1      | 7.9   | 0.99  | 4.8       | $[-\infty, 6.1 \times 10^{-2}]$  |
| 2      | -1.8  | 0.09  | 3.0       | $[-\infty, 1.0 \times 10^{-2}]$  |
| 3      | -3.6  | 0.02  | -1.2      | $[-\infty, -2.8 \times 10^{-2}]$ |
| 4      | -6.0  | 0.004 | -1.3      | $[-\infty, -6.8 \times 10^{-2}]$ |
| 5      | -5.6  | 0.005 | -2.8      | $[-\infty, -8.3 \times 10^{-2}]$ |

130  
 131 **Supplementary Table S2.** Full statistical results of the **Figure. 2G** in the main text. All statistical  
 132 results here are obtained by one-side t-tests.

| Layers | $t_{(3)}$ | p     | Cohen's d | 95% confidence interval          |
|--------|-----------|-------|-----------|----------------------------------|
| 1      | 3.9       | 0.99  | -2.6      | $[-\infty, 4.9 \times 10^{-1}]$  |
| 2      | -1.8      | 0.08  | -4.5      | $[-\infty, 9.6 \times 10^{-2}]$  |
| 3      | -3.5      | 0.02  | -4.2      | $[-\infty, -2.3 \times 10^{-1}]$ |
| 4      | -5.6      | 0.005 | -3.7      | $[-\infty, -5.6 \times 10^{-1}]$ |
| 5      | -4.7      | 0.008 | -3.1      | $[-\infty, -7.7 \times 10^{-1}]$ |

133  
 134 **Supplementary Table S3.** Full statistical results of the **Figure. 3F** in the main text. All statistical  
 135 results here are obtained by one-side t-tests.

| Layers | $t_{(3)}$ | p    | Cohen's d | 95% confidence interval |
|--------|-----------|------|-----------|-------------------------|
| 1      | 14        | 0.99 | 12        | $[-\infty, 0.47]$       |
| 2      | 7.2       | 0.99 | 3.8       | $[-\infty, 1.16]$       |
| 3      | 0.9       | 0.79 | 0.6       | $[-\infty, 3.53]$       |
| 4      | -0.4      | 0.36 | -0.25     | $[-\infty, 1.94]$       |
| 5      | -1.3      | 0.15 | -0.89     | $[-\infty, 1.04]$       |

136  
 137

138 **Supplementary Table S4.** Full statistical results of the **Figure. 3H** in the main text. All statistical  
 139 results here are obtained by one-side t-tests.

| Layers | $t_{(3)}$ | p     | Cohen's d | 95% confidence interval        |
|--------|-----------|-------|-----------|--------------------------------|
| 1      | 10.6      | <.001 | 2.0       | $[2.0 \times 10^{-3}, \infty]$ |
| 2      | 11.0      | <.001 | 8.7       | $[3.2 \times 10^{-3}, \infty]$ |
| 3      | 67.5      | <.001 | 47        | $[3.6 \times 10^{-1}, \infty]$ |
| 4      | 17.7      | <.001 | 14        | $[1.3 \times 10^{-1}, \infty]$ |
| 5      | 8.30      | .002  | 5.8       | $[1.5 \times 10^{-2}, \infty]$ |

140  
 141 **Supplementary Table S5.** Full statistical results of the **Figure. 4B** in the main text. All statistical  
 142 results here are obtained by one-side t-tests.

| Layers | Comparisons (indicated by different color bars) | $t_{(3)}$ | p     | Cohen's d | 95% confidence interval          |
|--------|---|-----------|-------|-----------|----------------------------------|
| 1      | Blue vs. yellow                                 | -18       | <.001 | -2.8      | $[-\infty, -1.6 \times 10^{-1}]$ |
|        | Green vs. Blue                                  | -18       | <.001 | -11       | $[-\infty, -1.1]$                |
|        | Cyan vs. Green                                  | 17        | <.001 | 7.9       | $[0.73, \infty]$                 |
| 2      | Blue vs. yellow                                 | -12       | <.001 | -3.9      | $[-\infty, -5.2 \times 10^{-1}]$ |
|        | Green vs. Blue                                  | -4.3      | .01   | -2.9      | $[-\infty, -7.5 \times 10^{-1}]$ |
|        | Cyan vs. Green                                  | 5.2       | .006  | 2.9       | $[0.91, \infty]$                 |
| 3      | Blue vs. yellow                                 | -22       | <.001 | -3.0      | $[-\infty, -8.5 \times 10^{-1}]$ |
|        | Green vs. Blue                                  | -16       | <.001 | -4.7      | $[-\infty, -1.2]$                |
|        | Cyan vs. Green                                  | 17        | <.001 | 7.4       | $[1.5, \infty]$                  |
| 4      | Blue vs. yellow                                 | -18       | <.001 | -2.5      | $[-\infty, -9.2 \times 10^{-1}]$ |
|        | Green vs. Blue                                  | -32       | <.001 | -2.3      | $[-\infty, -1.03]$               |
|        | Cyan vs. Green                                  | 7.2       | .002  | 3.9       | $[9.9 \times 10^{-1}, \infty]$   |
| 5      | Blue vs. yellow                                 | -23       | <.001 | -1.3      | $[-\infty, -7.8 \times 10^{-1}]$ |
|        | Green vs. Blue                                  | -9.5      | .001  | -2.4      | $[-\infty, -1.0]$                |
|        | Cyan vs. Green                                  | 17        | <.001 | 4.7       | $[1.5, \infty]$                  |

143  
 144 **Supplementary Table S6.** Full statistical results of the **Figure. 5C** in the main text. All statistical  
 145 results here are obtained by one-side t-tests.

| Layers | $t_{(3)}$ | p     | Cohen's d | 95% confidence interval          |
|--------|-----------|-------|-----------|----------------------------------|
| 1      | -1.2      | .16   | -0.54     | $[-\infty, -8.6 \times 10^{-4}]$ |
| 2      | -2.1      | .06   | -0.36     | $[-\infty, 2.9 \times 10^{-4}]$  |
| 3      | -3.1      | .03   | -0.36     | $[-\infty, -5.9 \times 10^{-4}]$ |
| 4      | -7.1      | .003  | -4.0      | $[-\infty, -1.3 \times 10^{-2}]$ |
| 5      | -9.3      | .001  | -5.7      | $[-\infty, -2.9 \times 10^{-2}]$ |
| 6      | -21       | <.001 | -11.7     | $[-\infty, -7.4 \times 10^{-2}]$ |

146

147 **Supplementary Table S7.** Full statistical results of the **Figure. 5D** in the main text. All statistical  
 148 results here are obtained by one-side t-tests.

| Layers | $t_{(3)}$ | p     | Cohen's d | 95% confidence interval          |
|--------|-----------|-------|-----------|----------------------------------|
| 1      | -2.4      | .05   | -1.7      | $[-\infty, -1.8 \times 10^{-4}]$ |
| 2      | -3.2      | .02   | -0.21     | $[-\infty, -1.4 \times 10^{-1}]$ |
| 3      | -1.0      | .19   | -0.12     | $[-\infty, 1.9 \times 10^{-1}]$  |
| 4      | -7.2      | .003  | -2.7      | $[-\infty, -1.4]$                |
| 5      | -13       | <.001 | -3.8      | $[-\infty, -1.5]$                |
| 6      | -26       | <.001 | -9.1      | $[-\infty, -2.7]$                |

149

150

151 **Supplementary Note 1: Additional explanations on the stepwise analyses**

152 We highlight two important aspects of such stepwise analyses. First, we use signal rotation and  
153 manifold warping to reflect changes in the direction of the signal vector and the shape of the  
154 covariance, which in theory can occur independently (Fig. 3I&K). But what matters for the fidelity  
155 of population representations is the relative geometric relationship between covariance and signal  
156 vector rather than their absolute magnitude [9,13](#) (Fig. 4C-F). Because learning changes both signal  
157 vector and covariance, we argue that their contributions to improved population representations  
158 should be evaluated together. In Fig. 4B, signal rotation and manifold warping appear to have  
159 strong positive and negative effects on information, respectively. However, their joint effect is  
160 modest (i.e., the magenta bars compared to the blue bars in Fig. 4B). Thus, we conclude that  
161 manifold shrinkage is the primary contributor to improved stimulus information. Second, the  
162 results are not affected by the order in which the mechanisms are applied changes. The effects of  
163 signal rotation and manifold warping should be evaluated together, but the effects of signal  
164 enhancement and manifold shrinkage are independent. Hence, changing the order does not affect  
165 our results as long as signal rotation and manifold warping are carried out consecutively (see  
166 Supplementary Fig. S5 for the order effect).

167

168 **Supplementary Note 2: Signal enhancement and manifold shrinkage in multivariate  
169 sensory neural populations and one-dimensional decision unit**

170 In this section, we theoretically derive the relationship between the analyses of population  
171 representations and the 1D responses of the decision unit. Assuming  $N$  sensory neurons in a  
172 population, we denote the mean firing rates towards the reference and the target stimuli as  $\vec{\mu}_{ref}$ ,  
173 and  $\vec{\mu}_{tgt}$ , their covariance as  $\Sigma_{ref}$  and  $\Sigma_{tgt}$ . We define the optimal linear decoder as  $\vec{w}$ . The  
174 decoder linearly transforms the high-dimensional population responses onto a decision plane to  
175 form the activity of the decision unit. The mean firing rates of the decision unit towards the two  
176 stimuli should be  $\vec{w} \bullet \vec{\mu}_{ref}$  and  $\vec{w} \bullet \vec{\mu}_{tgt}$  respectively while the variance should be  $\vec{w}^T \Sigma_{ref} \vec{w}$  and  
177  $\vec{w}^T \Sigma_{tgt} \vec{w}$ . According to the signal detection theory, we can define the separation between the two  
178 1-d decision distributions to be the difference between the classes over the variance within the  
179 classes as follows:

180 
$$J(\vec{w}) = \frac{(\vec{w} \bullet \vec{\mu}_{ref} - \vec{w} \bullet \vec{\mu}_{tgt})^2}{\vec{w}^T \Sigma_{ref} \vec{w} + \vec{w}^T \Sigma_{tgt} \vec{w}} = \frac{(\vec{w} \bullet (\vec{\mu}_{ref} - \vec{\mu}_{tgt}))^2}{\vec{w}^T (\Sigma_{ref} + \Sigma_{tgt}) \vec{w}} \quad (S1)$$

181 In a linear discriminant analysis, the objective is to find a  $\vec{w}$  that maximize  $J(\vec{w})$ , when  $\vec{w} =$   
 182  $k(\Sigma_{ref} + \Sigma_{tgt})^{-1}(\vec{\mu}_{ref} - \vec{\mu}_{tgt})$ . The Eq. S1 can be further written as follows:

$$183 J(\vec{w}) = \frac{(\vec{w} \bullet (\vec{\mu}_{ref} - \vec{\mu}_{tgt}))^2}{\vec{w}^T (\Sigma_{ref} + \Sigma_{tgt}) \vec{w}} = \frac{k(\vec{\mu}_{ref} - \vec{\mu}_{tgt})^T (\Sigma_{ref} + \Sigma_{tgt})^{-1} (\vec{\mu}_{ref} - \vec{\mu}_{tgt})}{\sqrt{k^2 (\vec{\mu}_{ref} - \vec{\mu}_{tgt})^T (\Sigma_{ref} + \Sigma_{tgt})^{-1} (\vec{\mu}_{ref} - \vec{\mu}_{tgt})}} = \frac{LFI}{\sqrt{LFI}}, \quad (S2)$$

184 Note that after the linear transformation, the numerator—the difference of the means (signal  
 185 separation defined in ref. 11)—is proportional to the linear Fisher information calculated from the  
 186 sensory neuron population. The denominator—the square root of the variance of the two decision  
 187 distributions (noise fluctuations defined in ref. 11)—is proportional to the squared root of the linear  
 188 Fisher information. Therefore, their definitions of signal and noise actually reflect all four possible  
 189 mechanisms identified in our main text.

190 More importantly, we also applied the similar analyses as in ref. 11 to the population  
 191 responses in layer 1 of the orientation DCNN. We replicated the results reported in ref. 11. We  
 192 also applied the same analyses to the voxel response data in the human fMRI experiment, and  
 193 replicated the Figure 6 in ref. 12. In summary, we analytically derive the relationship between the  
 194 signal/noise in high-dimensional population responses and the signal/noise in the one-dimensional  
 195 responses of the decision unit. The two seemingly opposite conclusions can coexist in both  
 196 artificial and biological visual systems.

197 **Supplementary Note 3: Computer simulations of perceptual consequences predicted by the**  
198 **neural geometry approach**

199 Although the neural geometry approach primarily addresses population response changes at the  
200 neural level, an intriguing question arises: what are the perceptual consequences of VPL as  
201 predicted by the four mechanisms? To illustrate the perceptual consequences predicted by our  
202 approach, we conducted systematic computer simulations to examine the effects of VPL of an  
203 orientation discrimination task on performance in an orientation estimation task, revealing  
204 perceptual orientation bias [14](#), and in a contrast detection task [15](#).

205 *Simulated effects of VPL of orientation discrimination on perceptual orientation bias.* It is  
206 well known that humans show biases in orientation estimation [14](#). Typically, estimated orientation  
207 is a linear function of true orientation, with perceptual bias characterized by the intercept and slope,  
208 as illustrated in Supplementary Fig. S7A. The intercept represents the degree of apparent  
209 orientation bias relative to the true orientation of a presented stimulus, while the slope represents  
210 the rate of change in orientation bias. We investigated how each mechanism (i.e., signal  
211 enhancement, manifold shrinkage, signal rotation, and manifold warping) induced by VPL in  
212 orientation discrimination influences the intercepts and slopes in the orientation estimation task  
213 [14](#).

214 As shown in the first and second rows of Supplementary Fig. S7B, the intercept generally  
215 decreases as manifold shrinkage and signal rotation become stronger, aligning with previous  
216 empirical findings [16](#). Signal enhancement also increases the slope, while manifold shrinkage  
217 initially decreases the slope and subsequently increases it. The observed increased slope after  
218 learning, reported in a previous study [16](#), is consistent with the effects of signal enhancement and  
219 partially with those of manifold shrinkage. In summary, the effects of VPL on perceptual  
220 orientation bias are consistently simulated by signal enhancement, manifold shrinkage, and/or  
221 signal rotation.

222 *Simulated effects of VPL of orientation discrimination on contrast detection.* As shown in  
223 the third row of Supplementary Fig. S7B, the simulation results indicate that signal enhancement,  
224 manifold shrinkage, and manifold warping induced by VPL of orientation discrimination reduce  
225 the contrast detection threshold. This accounts for the effect of VPL of orientation discrimination  
226 on contrast detection, aligning with existing experimental findings [15](#).

227        Simulation methods: we focused on the effects of orientation discrimination learning on  
 228 two perceptual tasks: orientation estimation and contrast detection. To achieve this, we began with  
 229 a simple neural population model that includes a pool of orientation-selective neurons. Each  
 230 neuron's tuning follows a von Mises function:

$$231 \quad f(\theta, \mu_i) = Ae^{\frac{\cos(2(\theta - \mu_i)) - 1}{2\sigma^2}} \quad (S3)$$

232 where  $A$  is the maximum response of this neuron,  $\sigma$  is the bandwidth of the tuning curve.  $\mu_i$  is  
 233 uniformly sampled from  $-90^\circ$  to  $90^\circ$ . A total of 360 neurons are simulated. Additionally, a  
 234 correlated gaussian noise with zero mean was added to each neuron.

$$235 \quad N_{gauss} \sim Normal(0, \sqrt{V}^T \rho \sqrt{V}) \quad (S4)$$

236 Here  $V$  is a diagonal matrix representing the variance of each neuron, with each element of  $V$  set  
 237 to 0.01.  $\rho$  represents the noise correlation between neurons. Since noise correlations are typically  
 238 positively correlated with signal correlations [17](#), we set the correlation  $\rho_{ij}$  between neuron  $i$  and  
 239 neuron  $j$  to be proportional to their signal correlation  $s_{ij}$ :

$$240 \quad \rho_{ij} = k s_{ij} \quad (S5)$$

241 Here  $k$  is set to be 0.1, and the signal correlation  $s_{ij}$  is defined as

$$242 \quad s_{ij} = \frac{\sum_\theta f(\theta, \mu_i)f(\theta, \mu_j)}{\sqrt{\sum_\theta f(\theta, \mu_i)^2 \sum_\theta f(\theta, \mu_j)^2}} \quad (S6)$$

243 We also added Poisson noise to each neuron. Thus, the final firing rate of the  $i$ -th neuron is:

$$244 \quad f_i = f(\theta, \mu_i) + N_{gauss,i} + Possion(f(\theta, \mu_i) + N_{gauss,i}) \quad (S7)$$

245 To examine the effects of each mechanism, we directly manipulated the tuning and noise  
 246 characteristics of the neurons according to the four mechanisms. The specific manipulations for  
 247 each mechanism were as follows:

- 248 • Signal enhancement: We increased the response amplitude  $A$  from 0.05 to 0.25, 0.5, 1.0.
- 249 • Manifold shrinkage: We reduced  $V$  from 0.01 to 0.09, 0.08, and 0.07.
- 250 • Signal rotation: We reduced  $\sigma$  from 0.5 to 0.4, 0.3, 0.2, and 0.1 to sharpen the tuning curves,  
 251 while simultaneously reducing  $A$  from 0.05 to approximately 0.043, 0.038, 0.03, and 0.023 to avoid  
 252 the signal enhancement effect.
- 253 • Manifold warping: We reduced  $k$  from 0.1 to 0.03, 0.01, and 0.005.

254 After inducing each mechanism in the neural population, we assessed their effects on  
255 psychophysical tasks by performing linear decoding of stimuli. To simulate perceptual illusions in  
256 the orientation estimation task, we used an ideal observer model and applied the optimal decoding  
257 weights as described in Supplementary Note 2. For the contrast detection task, we simulated  
258 behavioral performance by assuming a similar ideal observer who discriminates between stimulus  
259 and noise, rather than two stimuli. The code for these simulations is publicly available at  
260 [https://github.com/Yu-AngCheng/neural\\_geometry\\_VPL](https://github.com/Yu-AngCheng/neural_geometry_VPL).

261

## 262 **Supplementary Note 4: Supplementary methods for the human fMRI experiment**

263 *Experimental stimuli and procedures.* The random dot motion stimuli were presented to  
264 the participants on different display devices. In the behavioral sessions, a 40 cm wide CRT monitor  
265 with a resolution of  $1024 \times 768$  pixels and a refresh rate of 60 Hz was used. In addition, a 48 cm  
266 wide LCD projector with a resolution of  $1024 \times 768$  pixels and a refresh rate of 60 Hz was used  
267 to display the stimuli in the fMRI sessions. Stimuli were generated and displayed using  
268 Psychtoolbox 3.0 in the MATLAB environment. To ensure consistency, the viewing distance in  
269 the behavioral sessions was set at 75 cm from the screen. The method for generating the random  
270 dot stimuli followed the approach described in ref. 18. All stimuli were presented within an  
271 imperceptible aperture of  $10^\circ$  diameter centered on a black background. At any given moment, the  
272 aperture contained 400 dots, each moving in unison in the same direction at a speed of 4 %/s. Dots  
273 that moved out of the aperture were repositioned on the opposite side of the aperture to maintain a  
274 constant dot density.

275 At the beginning of each trial, a red reference cross was displayed for 500 ms. The  
276 orientation of the long arm of this red cross served as the reference direction for the subsequent  
277 motion stimulus. Following the reference cross, a red fixation dot appeared on the screen and  
278 remained there for the duration of the trial. After an unpredictable delay of 500 to 1000 ms, the  
279 motion stimulus was presented for 1500 ms or until participants responded. During the trial,  
280 participants were asked to indicate whether the motion stimulus rotated clockwise or  
281 counterclockwise relative to the orientation of the long arm of the reference cross. They were  
282 required to respond by pressing one of two designated keys according to their chosen direction  
283 preference.

284        The overall experimental design consisted of three distinct phases: the pre-test phase (2  
285 days), the training phase (10 days), and the post-test phase (2 days). Prior to the pre-test phase,  
286 participants underwent a practice session consisting of 80 trials with an angular difference of 8° to  
287 familiarize themselves with the task. Following this practice session, baseline performance was  
288 measured for each participant using a 4° angular difference. This baseline assessment was  
289 performed using four blocks of 30 trials each, with reference directions set at 45° and 135° (90°  
290 representing the vertical upward direction). The pre-test functional magnetic resonance imaging  
291 (fMRI) session was conducted on day 2. Each participant completed four runs of the direction  
292 discrimination task, with each run consisting of 60 task trials (30 trials for each direction of 45°  
293 and 135°) interspersed with 15 fixation trials. The order of these trials was pseudorandomized  
294 within each run and across participants, except that the first two trials and the last three trials in  
295 each run were designated as fixation trials. In addition, functional localizer runs were conducted  
296 within the same session (see ROI definitions).

297        The behavioral training phase lasted 10 days and was designed to ensure that participants  
298 reached ceiling performance. On each training day, participants completed 10 training runs, each  
299 run consisting of 60 task trials and 10 fixation trials. The total duration of each training session  
300 was approximately 1 hour. Training began with an initial angular difference of 4 °, and this angular  
301 difference remained fixed for each training session. After reaching a minimum accuracy of 79.4%  
302 on the preceding day, the angular difference for the following day was reduced to one of the  
303 predetermined options (3, 2, 1.5, or 1°) in order to increase task difficulty. This procedure was  
304 designed to keep participants engaged and challenge them with increasingly difficult trials. A fixed  
305 set of difficulty levels was used instead of staircase procedures to ensure a consistent level of task  
306 difficulty. This approach resulted in the creation of more challenging trials over time. Half of the  
307 participants were trained on 45° stimuli, while the other half on 135° stimuli. During training,  
308 auditory feedback was provided when participants made incorrect responses. In addition, visual  
309 feedback in the form of “slow down” or “hurry up” prompts was presented when the response time  
310 was faster than 250 ms or slower than 1500 ms, respectively. These feedback mechanisms were  
311 implemented to help participants refine their performance during the training process.

312        The procedure of the post-test phase was identical to the pre-test phase, except that the  
313 fMRI session was conducted before the behavioral session.

314        MRI data acquisition. The fMRI sessions were conducted on Day 2 at the pre-test phase  
315 and on Day 1 at the post-test phase. On Day 2 at pre-test, besides the four task runs, we also  
316 collected 1 run of T1 images, 1 run of motion localizer, 1-2 retinotopic runs. The retinotopic runs  
317 used the standard phase encoding method [19](#). The motion-responsive voxels (V3A and hMT+)  
318 were identified with a localizer procedure [20](#). We presented random dot stereograms, which were  
319 static for 24 s and then traveled toward and away from the fixation for 8 s. The moving/stationary  
320 cycle repeated nine times. The size of the stimulus aperture was the same as that used in the main  
321 experiment.

322

323 **Supplementary Note 5: Supplementary methods for the macaque multielectrode recording  
324 experiment**

325 Because we analyzed preprocessed monkey data, most of the experimental procedures have been  
326 published previously in detail. We provide more method details here only to avoid cross-  
327 referencing. To preserve as much of the method details as possible, we duplicate most of the  
328 noncritical text from the original paper [7](#) for reference purposes only. All duplicated text is  
329 italicized.

330        Surgical, headpost implementation, and electrode implementation procedures. We  
331 duplicated most of the text as italics from the original paper:

332 *Details of surgical procedures and postoperative care have been published elsewhere [21](#). An  
333 initial surgical operation was performed under sterile conditions, in which a custom-made head  
334 post (PEEK, Tecapeek) was embedded into a dental acrylic head stage.*

335        *For surgical preparation, animals were sedated with ketamine. During surgery, anesthesia  
336 and analgesia were maintained by sevoflurane (gaseous, 1–3%) and alfentanil (intravenous 156  
337 µg/kg/h), respectively. Blood pressure, rectal temperature, blood oxygen saturation and end tidal  
338 CO<sub>2</sub> were measured continuously. After the surgery, analgesic (metacam 0.1/kg) and prophylactic  
339 antibiotics (cephorex 0.5 ml/kg) was given for 3–5 days.*

340        *During surgery, the animals were placed in a stereotaxic head holder and the skull  
341 overlying the occipital and posterior temporal cortices was exposed. A craniotomy was made to  
342 remove the bone overlying V1, V2 and dorsal V4, using a pneumatic drill. The bone was kept in  
343 sterile 0.9% NaCl for refitting at the end of the surgery. The dura was opened to allow access to  
344 V4. Microelectrode chronic Utah arrays, attached to a CerePortTM base (Blackrock®*

345  *Microsystems, connection dimensions of 16.5 mm [height] × 19 mm [base diameter] × 11 mm*  
346 *[body diameter]), were implanted under sterile conditions in the cortex, using a Blackrock*  
347 *microarray inserter. In monkey 1, two 4 × 5 grids of microelectrodes were implanted in area V4;*  
348 *in monkey 2, a 5 × 5 grid was implanted in V4. Electrodes were 1 mm in length, and their tips*  
349 *reached depths of up to 1 mm. Wire bundles were held in place with biologically compatible glue*  
350 *(histoacrylic), and the connector (CerePortTM) was secured to the skull with titanium bone screws.*  
351 *Following array insertion, the Dura was re-sutured over the array, the exposed area was thinly*  
352 *covered with sterile Tisseel Lyo two-component fibrin sealant (Baxter Healthcare), and the bone*  
353 *flap was reinserted into the skull (before the Tisseel had fully set). The bone flap was cross bridged*  
354 *to the surrounding skull using Synthes orbital plate fragments and Synthes titanium bone screws.*

355 *The electrode arrays were inserted under visual control into the gyrus between the lunate*  
356 *sulcus and the superior temporal sulcus. The recording locations were confirmed to be in area V4*  
357 *in both animals via visual inspection immediately postmortem and by analysis of postmortem*  
358 *Nissl-stained brain sections.*

359 *Behavioral task and training procedures.* More detailed descriptions of task and training  
360 procedures are duplicated as italics from the original paper.

361 *Each monkey was trained in a contrast discrimination task in which he differentiated between two*  
362 *successively presented stationary Gabor gratings based on their relative contrasts (Fig. 7B).*

363 *Monkeys were initially trained on a very basic version of the contrast discrimination task*  
364 *at a location in the upper visual field, i.e., at a substantial distance from the RFs covered by our*  
365 *electrodes, which were located in the lower left visual field (for details, see below). When the*  
366 *animal understood the main concept of the task in the upper visual field, the stimuli were shifted*  
367 *to the left lower visual field. The stimuli (Gabor gratings,  $\sigma = 4^\circ$ , spatial frequency = 2 cycle per*  
368 *degree, orientation = 90° vertical) were initially presented at an azimuth of  $-5^\circ$  and an elevation*  
369 *of  $-16^\circ$  in both monkeys (left and bottom compared to the fixation point). These coordinates*  
370 *covered the V4 RFs.*

371 *Each trial was initiated when the monkey held a touch bar and fixated on a small fixation*  
372 *spot (diameter =  $0.1^\circ$ , fixation window =  $2^\circ \times 2^\circ$ ) which was presented on a grey background*  
373 *( $52.17 \text{ cd/m}^2$ ). During the trial, if the monkey broke fixation before saccade cue onset or failed to*  
374 *respond within 1000 ms of the onset of the saccade cue, the trial was terminated immediately and*  
375 *followed by a 0.2 s timeout. We used different inter-stimulus intervals in the two animals for the*

376 following reason. We started training and recording in monkey 1, before doing so in monkey 2.  
377 We initially reasoned that a variable test onset would increase the animal's focus and thereby  
378 possibly learning. In monkey 2, the variable onset during the very basic training resulted in too  
379 many early trial abortions, which quickly vanished when we used a fixed delay. We therefore  
380 decided to use a fixed delay in that animal.

381 After monkeys performed well in this easy version of the task, the number of test contrasts  
382 was increased to 8 (5, 10, 20, 25, 35, 40, 60 and 90% contrast, on day 1 of the proper contrast  
383 discrimination task), then to 12 (10, 15, 20, 25, 27, 29, 31, 33, 35, 40, 50 and 60% contrast, on  
384 day 2 of the proper contrast discrimination task) and to 14 (10, 15, 20, 25, 27, 28, 29, 31, 32, 33,  
385 35, 40, 50 or 60% contrast, from day 3 of the proper contrast discrimination task). In order to  
386 motivate subjects to complete each trial and discourage them from guessing on difficult trials,  
387 stimulus drumming was carried out using the 'repetition with delay' function on CORTEX  
388 following error trials, i.e. enforcing the repeated presentation of a stimulus condition, until a  
389 minimum number of correct trials is accrued. Recording began simultaneously with the first day  
390 of training on the proper contrast discrimination task, but data analysis for the purpose of this  
391 paper was only performed from day 3 onwards, as this was the start of presenting the full range  
392 of contrasts.

393 Data acquisition and processing. Data acquisition and preprocess details are also not  
394 critical for this study. We duplicated the texts as italics.

395 Raw data were acquired at a sampling frequency of 32,556 Hz with a 24-bit analogue-to-digital  
396 converter, with minimum and maximum input ranges of 11 and 136,986 microvolts, respectively  
397 (pre-set by Neuralynx, Inc.), a DMA buffer count of 128 and a DMA buffer size of 10 ms, using a  
398 64-channel Digital Lynx 16SX Data Acquisition System (Neuralynx, Inc.). Digital referencing of  
399 voltage signals was performed prior to the recording of raw data, using commercially provided  
400 Cheetah 5 Data Acquisition Software v. 5.4.0 (Neuralynx, Inc.), to yield good SNRs for each  
401 channel.

402 Following each recording session, the raw data were processed offline using both  
403 commercial (Neuralynx, Inc.) and custom-written (Matlab, Mathworks) software. Signals were  
404 extracted using the Cheetah 5 Data Acquisition Software. The sampling frequency remained the  
405 same (32,556 Hz), while the bandpass filter frequency and the input range settings were  
406 individually tailored to each channel. Raw data were bandpass filtered with a low cut frequency

407 of 600 Hz and a high cut frequency of 4000 Hz and saved at 16-bit resolution. This stage of  
408 processing generated ‘continuous multi-unit activity (MUA)’ data, which was further processed  
409 to yield ‘spiking MUA’.

410 An iterative procedure was carried out on the continuous MUA signals for each channel,  
411 in which the threshold for spike extraction was varied according to a staircase procedure, in order  
412 to yield levels of spontaneous spiking MUA (before the onset of the sample stimulus) that were  
413 similar (within 1% of a ‘target’ level) across sessions. To set the target level for each channel, the  
414 threshold was initially selected manually for all channels and sessions, and a ‘representative’  
415 session was selected for each channel (i.e., a session with an ‘average’ SNR [see below for  
416 description] for that channel). Hence, the extraction of spiking MUA was performed such that  
417 spontaneous activity levels were standardized across recording sessions. As spontaneous activity  
418 levels were deliberately kept uniform across training days, we did (or could) not study whether  
419 spontaneous activity levels changed during training. What this method did allow, however, was  
420 the rigorous comparison of levels of stimulus-evoked activity across the training period, relative  
421 to spontaneous levels.

422 Receptive fields were mapped using a reverse correlation procedure 22 for each recording  
423 channel prior to training and recording. Additionally, orientation and spatial frequency tuning  
424 was determined using a reverse correlation procedure 22. RF locations and tuning preferences  
425 were highly consistent across the training period as determined by regular remapping while  
426 learning commenced (every 3–5 days).

427 Data exclusion. We used the exactly same methods to compute SNR of individual channels  
428 and applied the same criteria to select high profile channels. The SNR computation are described  
429 as below in the original paper:

430 The SNR was calculated for each channel on each day. The SNR was calculated as:

$$431 \quad SNR = \frac{Mean_{stimulus\ activity} - Mean_{spontaneous}}{SD_{spontaneous}} \quad (S8)$$

432 whereby the mean stimulus activity was obtained from 150 to 250 ms after test onset, while the  
433 mean spontaneous activity was obtained during the 300-ms period before test onset. SD is the  
434 standard deviation of the mean response. This was calculated for each test contrast condition,  
435 yielding 14 SNR values per recording session for a given channel. Trials were included regardless  
436 of whether the subject’s response was correct. The size of the SNR varied depending on the test

437 contrast. The highest of the 14 SNR values was then taken as being representative of the signal  
438 quality from a given channel for each session. Channels were included in the individual channel  
439 analyses if they had daily  $\text{SNR} \geq 1$ , on at least 80% of the total number of recording days.

440 Note that according to this criteria, 29 and 20 channels were included for monkey 1 and 2,  
441 respectively. Two channels in monkey 1 were found to show monotonically decreasing contrast  
442 response functions. We thus excluded the two channels for further analyses.

443 Population activity analyses. Previous studies have suggested that the variability of  
444 neuronal responses can be decomposed into two components—the trial-by-trial variability of firing  
445 rate and the intrinsic Poisson variability of spike generation. We hypothesize that VPL, as a top-  
446 down task signal, mainly manifests as the changes in the trial-by-trial variability of firing rate  
447 rather than altering the intrinsic Poisson statistics of the spike because the latter component is more  
448 driven by low-level biophysical properties of neuronal spikes. We used a simple multivariate  
449 Poisson-lognormal (MPLN) model as used in previous studies [23-26](#), which assumes that the spike  
450 counts (SC) of individual channels are generated by a homogeneous Poisson process, given the  
451 underlying firing rate ( $\vec{fr}$ ). The model also assumes that the underlying firing rate of channels is  
452 represented by a multivariate log-normal distribution with a mean vector  $\mu$  and covariance  $\Sigma$ ,  
453 which accounts for the covariability among channels and prevents negative firing rates. The model  
454 is expressed mathematically as:

$$455 \quad SC_i \sim \text{Poisson}(T \times fr_i) \quad (\text{S9})$$

$$456 \quad \vec{fr} \sim \text{log normal } (\vec{\mu}, \Sigma) \quad (\text{S10})$$

457 To infer the latent parameters  $\mu$  and  $\Sigma$ , we used the variational inference method with an Adam  
458 optimizer and a learning rate of 0.1. The prior of  $\mu$  was set to be Gaussian distributed with a mean  
459 of 1.5 and a standard deviation of 3. The prior of  $\Sigma$  was set through its Cholesky matrix, which is  
460 set to be Lewandowski-Kurowicka-Joe (LKJ) distributed with an  $\eta$  of 1 and standard deviation  
461 distribution to be an exponential distribution with a  $\lambda$  of 3. The mean of the posterior inference via  
462 variational inference was taken as our point estimate of  $\mu$  and  $\Sigma$ . We then defined the signal vector,  
463 signal separation, response variance, signal rotation angle, PC strength, and PC rotation angle, in  
464 a manner similar to our previous work with DCNN models and fMRI data. The only difference is  
465 that we now applied these calculations to multi-unit activity within a latent space characterized by  
466  $\mu$  and  $\Sigma$ . This was followed by the application of our established stepwise analysis approach.

467

## 468 REFERENCES

- 469 1 Hennequin, G., Ahmadian, Y., Rubin, D. B., Lengyel, M. & Miller, K. D. The Dynamical  
470 Regime of Sensory Cortex: Stable Dynamics around a Single Stimulus-Tuned Attractor  
471 Account for Patterns of Noise Variability. *Neuron*. **98**(4): 846-860 e845 (2018).
- 472 2 van Bergen, R. S. & Jehee, J. F. M. Modeling correlated noise is necessary to decode  
473 uncertainty. *Neuroimage*. **180**(Pt A): 78-87 (2018).
- 474 3 Cheng, Z. J., Yang, L., Zhang, W. H. & Zhang, R. Y. Representational Geometries Reveal  
475 Differential Effects of Response Correlations on Population Codes in Neurophysiology and  
476 Functional Magnetic Resonance Imaging. *J Neurosci*. **43**(24): 4498-4512 (2023).
- 477 4 Ecker, A. S. *et al.* Decorrelated neuronal firing in cortical microcircuits. *Science*.  
478 **327**(5965): 584-587 (2010).
- 479 5 Ruff, D. A. & Cohen, M. R. Attention can either increase or decrease spike count  
480 correlations in visual cortex. *Nat Neurosci*. **17**(11): 1591-1597 (2014).
- 481 6 Bejjanki, V. R., Beck, J. M., Lu, Z. L. & Pouget, A. Perceptual learning as improved  
482 probabilistic inference in early sensory areas. *Nat Neurosci*. **14**(5): 642-648 (2011).
- 483 7 Sanayei, M. *et al.* Perceptual learning of fine contrast discrimination changes neuronal  
484 tuning and population coding in macaque V4. *Nat Commun*. **9**(1): 4238 (2018).
- 485 8 Zhang, R. Y., Wei, X. X. & Kay, K. Understanding multivariate brain activity: Evaluating  
486 the effect of voxelwise noise correlations on population codes in functional magnetic  
487 resonance imaging. *PLOS Computational Biology*. **16**(8): e1008153 (2020).
- 488 9 Kohn, A., Coen-Cagli, R., Kanitscheider, I. & Pouget, A. Correlations and Neuronal  
489 Population Information. *Annu Rev Neurosci*. **39**:237-256 (2016).
- 490 10 Kanitscheider, I., Coen-Cagli, R., Kohn, A. & Pouget, A. Measuring Fisher information  
491 accurately in correlated neural populations. *PLOS Computational Biology*. **11**(6):  
492 e1004218 (2015).
- 493 11 Yan, Y. *et al.* Perceptual training continuously refines neuronal population codes in  
494 primary visual cortex. *Nat Neurosci*. **17**(10): 1380-1387 (2014).
- 495 12 Chen, N., Cai, P., Zhou, T., Thompson, B. & Fang, F. Perceptual learning modifies the  
496 functional specializations of visual cortical areas. *Proc Natl Acad Sci U S A*. **113**(20): 5724-  
497 5729 (2016).
- 498 13 Moreno-Bote, R. *et al.* Information-limiting correlations. *Nat Neurosci*. **17**(10): 1410-1417  
499 (2014).
- 500 14 Jazayeri, M. & Movshon, J. A. A new perceptual illusion reveals mechanisms of sensory  
501 decoding. *Nature*. **446**(7138): 912-915 (2007).
- 502 15 Levi, A., Shaked, D., Tadin, D. & Huxlin, K. R. Is improved contrast sensitivity a natural  
503 consequence of visual training? *J Vis*. **15**(10): 4 (2015).
- 504 16 Jia, K. & Li, S. Motion direction discrimination training reduces perceived motion  
505 repulsion. *Atten Percept Psychophys*. **79**(3): 878-887 (2017).
- 506 17 Averbeck, B. B., Latham, P. E. & Pouget, A. Neural correlations, population coding and  
507 computation. *Nat Rev Neurosci*. **7**(5): 358-366 (2006).
- 508 18 Britten, K. H., Shadlen, M. N., Newsome, W. T. & Movshon, J. A. The analysis of visual  
509 motion: a comparison of neuronal and psychophysical performance. *J Neurosci*. **12**(12):  
510 4745-4765 (1992).
- 511 19 Engel, S. A., Glover, G. H. & Wandell, B. A. Retinotopic organization in human visual  
512 cortex and the spatial precision of functional MRI. *Cereb Cortex*. **7**(2): 181-192 (1997).

- 513 20 Huk, A. C., Dougherty, R. F. & Heeger, D. J. Retinotopy and functional subdivision of  
514 human areas MT and MST. *J Neurosci.* **22**(16): 7195-7205 (2002).
- 515 21 Thiele, A., Delicato, L. S., Roberts, M. J. & Gieselmann, M. A. A novel electrode-pipette  
516 design for simultaneous recording of extracellular spikes and iontophoretic drug  
517 application in awake behaving monkeys. *J Neurosci Methods.* **158**(2): 207-211 (2006).
- 518 22 Gieselmann, M. A. & Thiele, A. Comparison of spatial integration and surround  
519 suppression characteristics in spiking activity and the local field potential in macaque V1.  
520 *Eur J Neurosci.* **28**(3): 447-459 (2008).
- 521 23 Goris, R. L., Movshon, J. A. & Simoncelli, E. P. Partitioning neuronal variability. *Nat  
522 Neurosci.* **17**(6): 858-865 (2014).
- 523 24 Henaff, O. J. *et al.* Primary visual cortex straightens natural video trajectories. *Nat  
524 Commun.* **12**(1): 5982 (2021).
- 525 25 Rabinowitz, N. C., Goris, R. L., Cohen, M. & Simoncelli, E. P. Attention stabilizes the  
526 shared gain of V4 populations. *Elife.* **4e08998** (2015).
- 527 26 Zhu, R. J. B. & Wei, X. X. Unsupervised approach to decomposing neural tuning  
528 variability. *Nat Commun.* **14**(1): 2298 (2023).
- 529

Corresponding author(s): Ru-Yuan Zhang

Last updated by author(s): Feb 1, 2025

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection CORTEX software (last updated 2013, <http://dally.nimh.nih.gov/index.html>) and Psychtoolbox 3.0 in MATLAB2013A was used to collect monkey neurophysiology and human fMRI data.

Data analysis For training deep neural networks, we used a Python 3.10.9 conda environment, including pytorch 1.13.1, scikit-learn 1.2.0. Details of full conda environment are provided at [https://github.com/Yu-AngCheng/neural\\_geometry\\_VPL](https://github.com/Yu-AngCheng/neural_geometry_VPL). For MRI preprocessing, we use SPM 12 and Brain Voyager QX (version 2.8.0). GLMdenoise package (version 1.4, <http://www.kendrickkay.net/GLMdenoise/>) was used to perform single trial general linear modeling. For monkey neurophysiology analysis, we use custom python code provided at [https://github.com/Yu-AngCheng/neural\\_geometry\\_VPL](https://github.com/Yu-AngCheng/neural_geometry_VPL).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All data to reproduce the figures in the main and supplementary information can be found at [https://github.com/Yu-AngCheng/neural\\_geometry\\_VPL](https://github.com/Yu-AngCheng/neural_geometry_VPL). The raw fMRI data and the raw monkey neurophysiological data are available upon reasonable request, which is consistent with the original published studies.

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

### Reporting on sex and gender

We collected information about the (biological) sex of the participants (self-reported), and kept the gender balance of the subjects roughly Twenty-two (12 females and 10 males) with no history of neurological or psychiatric disorders were recruited in the human fMRI experiment. We did not make any analysis of the gender of the subjects since this was beyond the scope of our research.

### Reporting on race, ethnicity, or other socially relevant groupings

Participants were not classified into different race, ethnicity or other social categories.

### Population characteristics

A total of 22 human subjects (10 males and 12 females, ages 17-25) participated in the experiment. All participants had normal or correct-to-normal vision. None of the participants were aware of the study's objectives.

### Recruitment

We recruited participants through an online advertisement placed on a university campus bulletin board system (BBS). They were primarily undergraduate or graduate students from the universities. There was no self-selection bias.

### Ethics oversight

The study obtained approval from the local ethics committee at Peking University (Protocol#: 2013-02-09).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

Behavioural & social sciences

Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

### Study description

This quantitative study examines whether tuning curve changes and/or noise correlation changes occur in the human brain over the course of visual training, and whether these two factors actually contribute to the improvements of population representations over the course of visual training.

### Research sample

A total of 22 human subjects (10 males and 12 females, ages 17-25) participated in the experiment. All participants had normal or correct-to-normal vision. None of the participants were aware of the study's objectives. The study was previously published in Jia, K. et al. (2018). Therefore, the 22 participants are representative in terms of visual perception.

### Sampling strategy

A convenience sampling is adopted, meaning the participants volunteered to attend the current study after seeing the advertisement. Sample size are chosen as comparable to previous visual perceptual learning studies using fMRI. The data were previously published in Jia, K. et al. (2018).

### Data collection

The random dot motion stimuli were presented to the participants on different display devices. In the behavioral sessions, a 40 cm wide CRT monitor with a resolution of 1024 × 768 pixels and a refresh rate of 60 Hz was used. In addition, a 48 cm wide LCD projector with a resolution of 1024 × 768 pixels and a refresh rate of 60 Hz was used to display the stimuli in the fMRI sessions. In all behavioral experiments, the participants were accompanied only by the researchers. In the fMRI experiments, the participants were accompanied only by the researchers and extra MRI technicians. A single-blind design was employed, i.e., participants were unaware of the purpose of this study.

### Timing

2013.01-2013.11

|                   |  |
|-------------------|--|
| Data exclusions   | Three participants in the fMRI were excluded due to drop-out and excessive head motion which was established prior to data collection. This exclusion criteria is common in fMRI research. |
| Non-participation | One participant dropped.   |
| Randomization     | We only recruit one group of 22 subjects and therefore no group randomization was performed. However, trials of different conditions are randomized.                                       |

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

|                                     |                               |
|-------------------------------------|-------------------------------|
| n/a                                 | Involved in the study         |
| <input checked="" type="checkbox"/> | Antibodies                    |
| <input checked="" type="checkbox"/> | Eukaryotic cell lines         |
| <input checked="" type="checkbox"/> | Palaeontology and archaeology |
| <input type="checkbox"/>            | Animals and other organisms   |
| <input checked="" type="checkbox"/> | Clinical data                 |
| <input checked="" type="checkbox"/> | Dual use research of concern  |
| <input checked="" type="checkbox"/> | Plants                        |

### Methods

|                                     |                        |
|-------------------------------------|------------------------|
| n/a                                 | Involved in the study  |
| <input checked="" type="checkbox"/> | ChIP-seq               |
| <input checked="" type="checkbox"/> | Flow cytometry         |
| <input type="checkbox"/>            | MRI-based neuroimaging |

## Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

### Laboratory animals

Two male macaque monkeys (5 and 14 years of age) were used in this study.

### Wild animals

The study did not involve wild animals

### Reporting on sex

Sex was not considered as a factor in the current study.

### Field-collected samples

The study did not contain samples collected from the field

### Ethics oversight

All procedures were approved by the Newcastle University Animal Welfare Ethical Review Board and carried out in accordance with the European Communities Council Directive RL 2010/63/EC, the US National Institutes of Health Guidelines for the Care and Use of Animals for Experimental Procedures and the UK Animals Scientific Procedures Act.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Plants

### Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

### Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

### Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.

## Magnetic resonance imaging

### Experimental design

#### Design type

Task state. Event-related design.

#### Design specifications

Data was first published in Jia et al. 2018, and was reanalyzed for this paper. In each fMRI session, subjects completed

|   |  |
|---|--|
| Design specifications                       | four runs of the motion direction discrimination task. Each run contained 30 trials for 45° and 135° (i.e., a total of 120 trials for each direction). Each run also contained 15 fixation trials and the trial order was randomized. In addition to the four runs of the motion direction discrimination task, we also collected one or two retinotopic mapping runs and a motion localizer run to define regions-of-interest.  |
| Behavioral performance measures             | Subjects were tested on direction discrimination around 45° and 135° (angular difference 40°, 120 trials for each direction) to assess their accuracy.   |
| <b>Acquisition</b>                          |  |
| Imaging type(s)                             | functional   |
| Field strength                              | All MRI data were acquired using a 12-channel phase array coil on a Siemens Trio 3T scanner at Peking University.  |
| Sequence & imaging parameters               | The T1-weighted anatomical data with a resolution of 1 × 1 × 1 mm <sup>3</sup> were collected for each subject. Echo-planar imaging (EPI) functional data were collected for the motion direction discrimination task, retinotopic mapping, and motion localizer experiments. EPI data were acquired using gradient echo-pulse sequences from 33 axial slices, covering the whole brain. The standard EPI sequence used for data acquisition was as follows: a repetition time of 2000 ms, an echo time of 30 ms, a flip angle of 90°, and a resolution of 3 × 3 × 3 mm <sup>3</sup> . The slice order was interleaved ascending.  |
| Area of acquisition                         | Whole brain  |
| Diffusion MRI                               | <input type="checkbox"/> Used <input checked="" type="checkbox"/> Not used   |
| <b>Preprocessing</b>                        |  |
| Preprocessing software                      | In Brain Voyager QX(version 2.8.0), the anatomical data were transformed into the Talairach coordinate space. For all functional data, the first four volumes of each functional run were discarded to allow the longitudinal magnetization to reach a steady state. The functional data underwent several standard preprocessing procedures, including slice timing correction, head motion correction, spatial smoothing, temporal high-pass filtering (GLM with Fourier basis set at 2 cycles), and linear trend removal. Brain Voyager QX (version 2.8.0) was also used to preprocess the data of the retinotopic mapping experiment and the motion localizer experiment. We used the standard phase-encoding method to define the retinotopic visual areas V1, V2, V3, and V3A 44,45. A generalized linear model (GLM) was then applied to the motion localizer data to define the motion-selective voxels (hMT+ and motion-selective voxels in IPS). The functional data of the motion direction discrimination task were preprocessed using SPM12 ( <a href="http://www.fil.ion.ucl.ac.uk/spm">www.fil.ion.ucl.ac.uk/spm</a> ). The data were aligned to the first volume of the first run of the first session, corrected for acquisition delay, and then normalized to the MNI coordinate space using an EPI template. We used the GLMdenoise package developed in ref. 46 without evoking multirun denoise procedures to estimate the single trial activity of voxels. |
| Normalization                               | The data were aligned to the first volume of the first run of the first session, corrected for acquisition delay, and then normalized to the MNI coordinate space using an EPI template.   |
| Normalization template                      | The functional data were normalized to the MNI coordinate space using an EPI template.   |
| Noise and artifact removal                  | The functional data underwent several standard preprocessing procedures, including slice timing correction, head motion correction, spatial smoothing, temporal high-pass filtering (GLM with Fourier basis set at 2 cycles), and linear trend removal.  |
| Volume censoring                            | For all functional data, the first four volumes of each functional run were discarded to allow the longitudinal magnetization to reach a steady state.   |
| <b>Statistical modeling &amp; inference</b> |  |
| Model type and settings                     | GLMdenoise package () was used without evoking multirun denoise procedures to estimate the single trial activity of voxels.  |
| Effect(s) tested                            | Effected were test before and after visual perceptual learning   |
| Specify type of analysis:                   | <input type="checkbox"/> Whole brain <input checked="" type="checkbox"/> ROI-based <input type="checkbox"/> Both   |
| Anatomical location(s)                      | We used functional ROIs, identified and tested using retinotopic mapping runs and motion localizer runs.   |
| Statistic type for inference                | cluster-wise (ROI-wise) analysis were applied  |
| (See <a href="#">Eklund et al. 2016</a> )   |  |
| Correction                                  | We did not run whole brain search analysis. Therefore, no correction were applied since we only focused on V3a and hMT+  |

## Models & analysis

n/a Involved in the study

Functional and/or effective connectivity

Graph analysis

Multivariate modeling or predictive analysis

Multivariate modeling and predictive analysis

Single trial activity of voxels within the ROIs were used to predict the conditions of the experiment (45° and 135°). 60 voxels were used for each ROI. A leave-one-trial-out cross validation was applied. The average accuracy over the test trials was used as the metric.