

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

In the format provided by the
authors and unedited

Supplementary Information for

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

Table of Content

Supplementary Figure S1: Neural consequences of orientation VPL in all five layers.

Supplementary Figure S2: The orientation DCNN replicates the neuronal response properties in the primate early visual cortex.

Supplementary Figure S3: Neural consequences of motion VPL in all six layers.

Supplementary Figure S4: Neural consequences of contrast discrimination learning in two monkeys.

Supplementary Figure S5: The effects of the mechanism order in the stepwise analyses.

Supplementary Figure S6: Replications of macaque neurophysiological and human fMRI experiments on signal enhancement.

Supplementary Figure S7: Predicted effects of orientation discrimination learning on estimation and detection tasks.

Supplementary Table S1-S7: Full statistical results of Figure 2F&G, Figure 3F&H, Figure 4B, and Figure 5C&D in the main text.

Supplementary Note 1: Additional explanations on the stepwise analyses

Supplementary Note 2: Analytical relation between the high-dimensional population responses and one-dimensional decision unit responses.

Supplementary Note 3: Computer simulations of perceptual consequences predicted by the neural geometry approach.

Supplementary Note 4: Supplementary methods for the human fMRI experiment.

Supplementary Note 5: Supplementary methods for the macaque multielectrode recording experiment.

Supplementary Figure S1

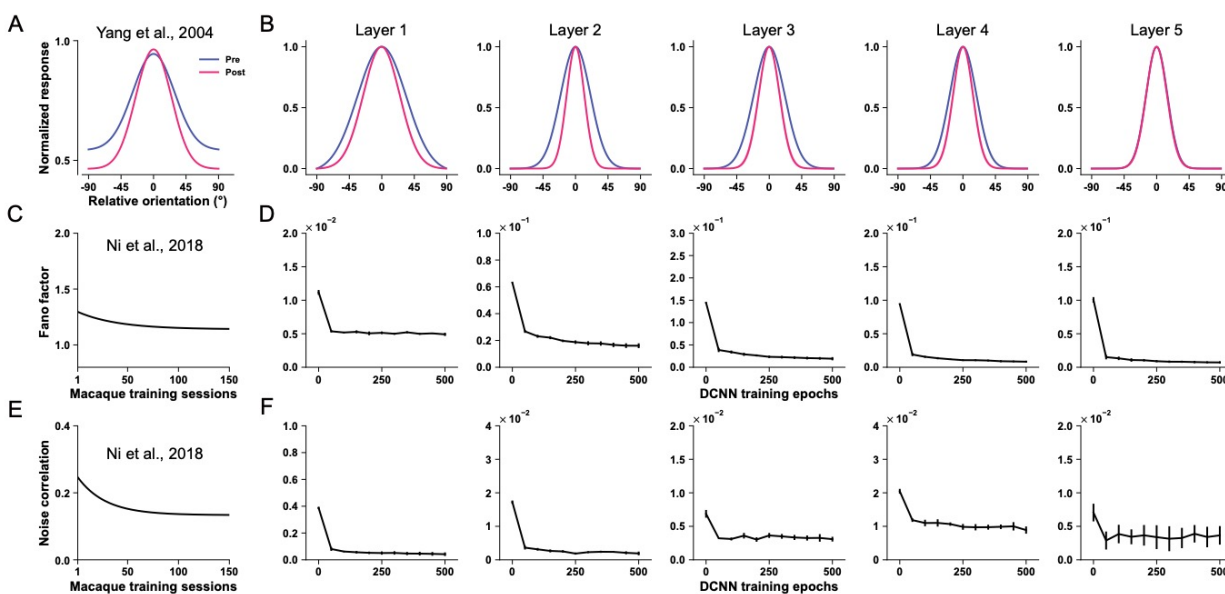


Figure S1. Tuning curve, Fano factor, and noise correlation changes in all five layers of the orientation DCNN. Perceptual training on the DCNN reduces the width of the tuning curves in layers 1-4. Perceptual training also reduces Fano factor and noise correlations in all five layers. Data are presented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

Supplementary Figure S2

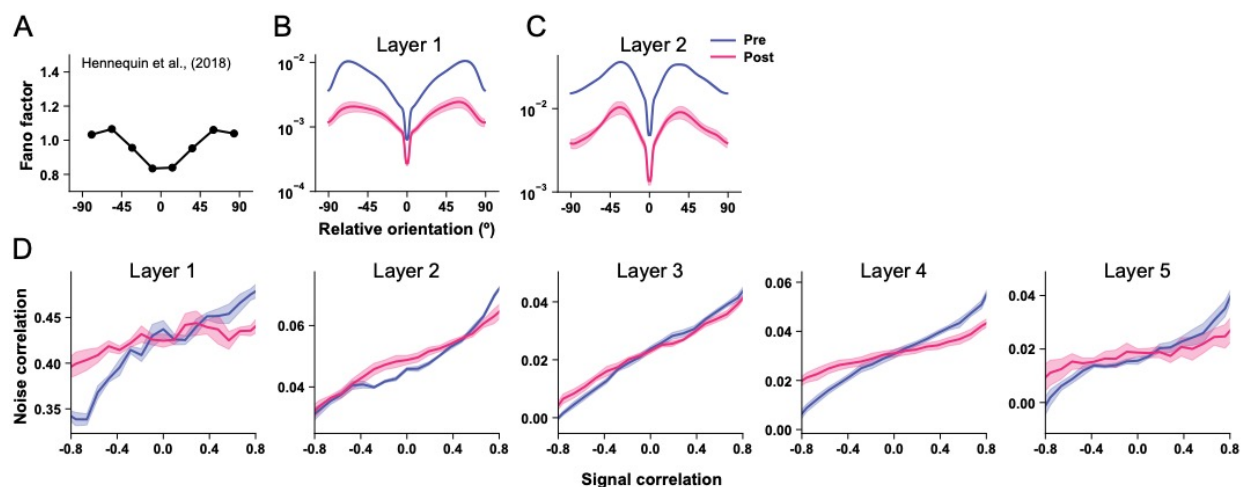


Figure S2. The DCNN of orientation VPL qualitatively replicates the relationship between orientation preferences and Fano factor in macaque V1 (A, reproduced from the datapoints from ref. 1). Here, the results of layers 1(B)&2(C) are plotted. The network also reproduces the positive relationship between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

Supplementary Figure S3

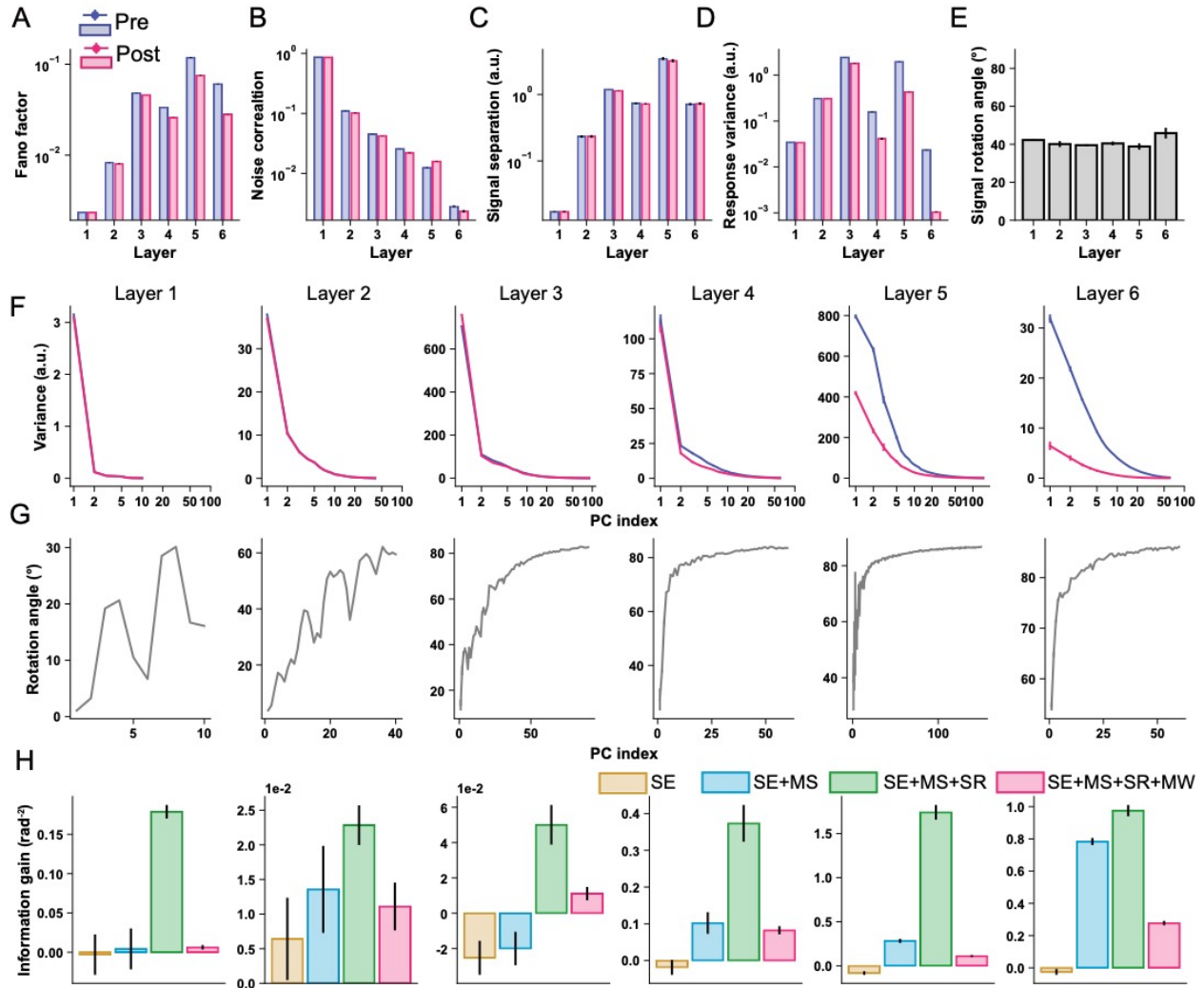


Figure S3. The results of all layers in the motion DCNN. The effects of motion direction discrimination training on Fano factor (A), noise correlations (B), signal separation (C), response variance (D), and signal rotation angle (E) in the DCNN model. The effects of training induced manifold warping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

66 **Supplementary Figure S4**

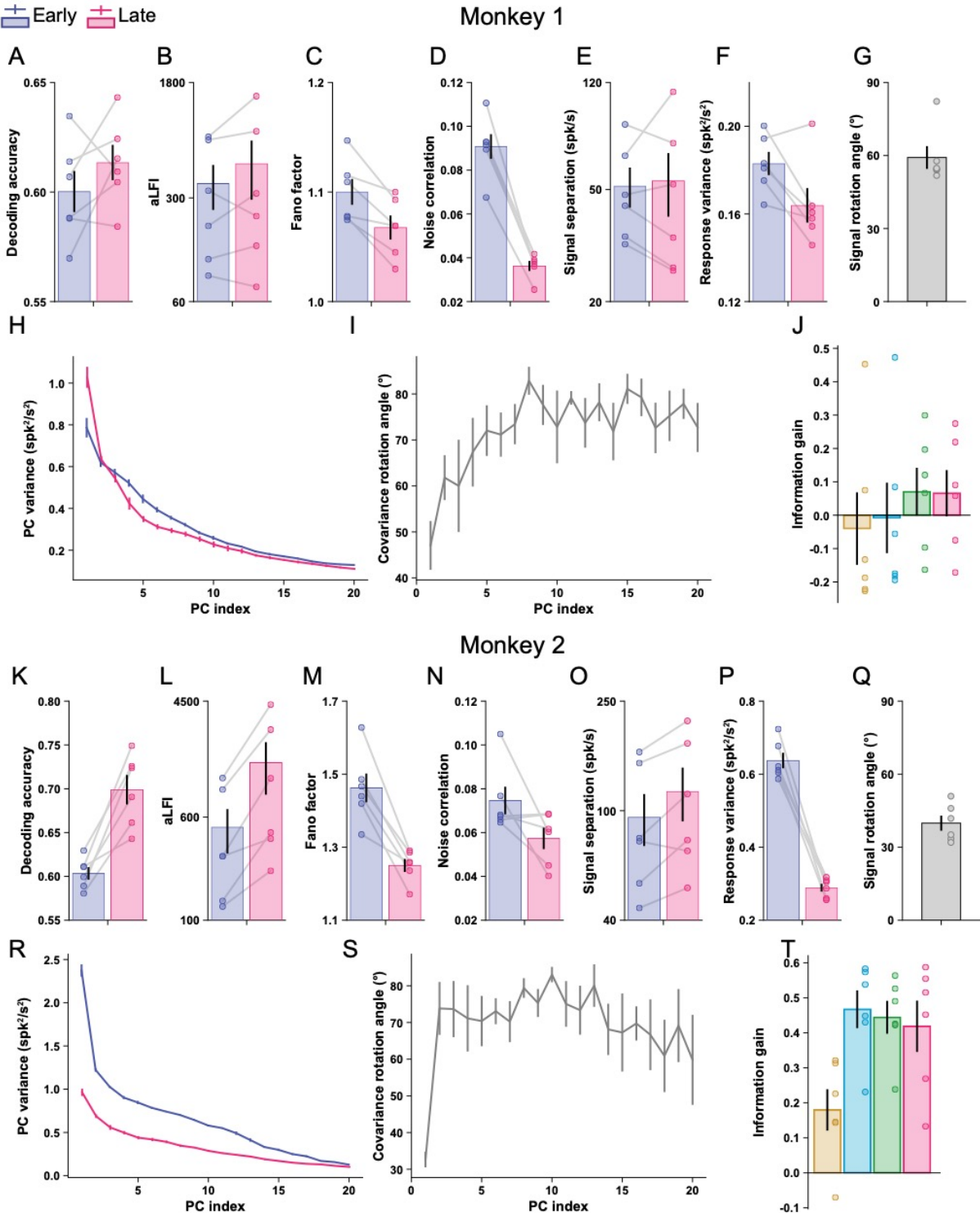


Figure S4. Population activity analyses for the two monkeys. The organization and conventions here are the same as Fig. 8 in the main text, except that we split the results for the two monkeys. Data are presented as mean values \pm S.E.M., with error bars indicate the S.E.M. across the six conditions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

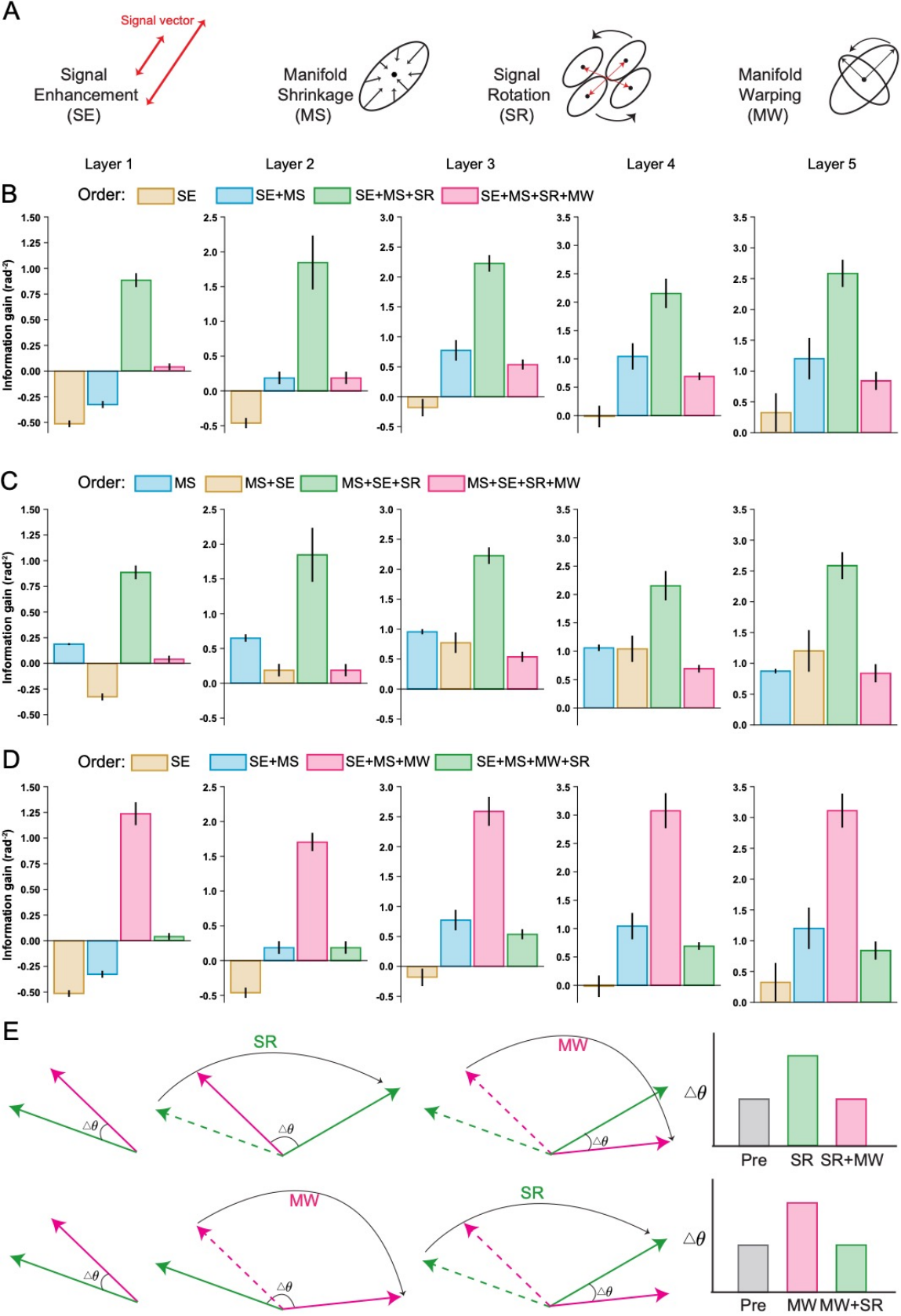
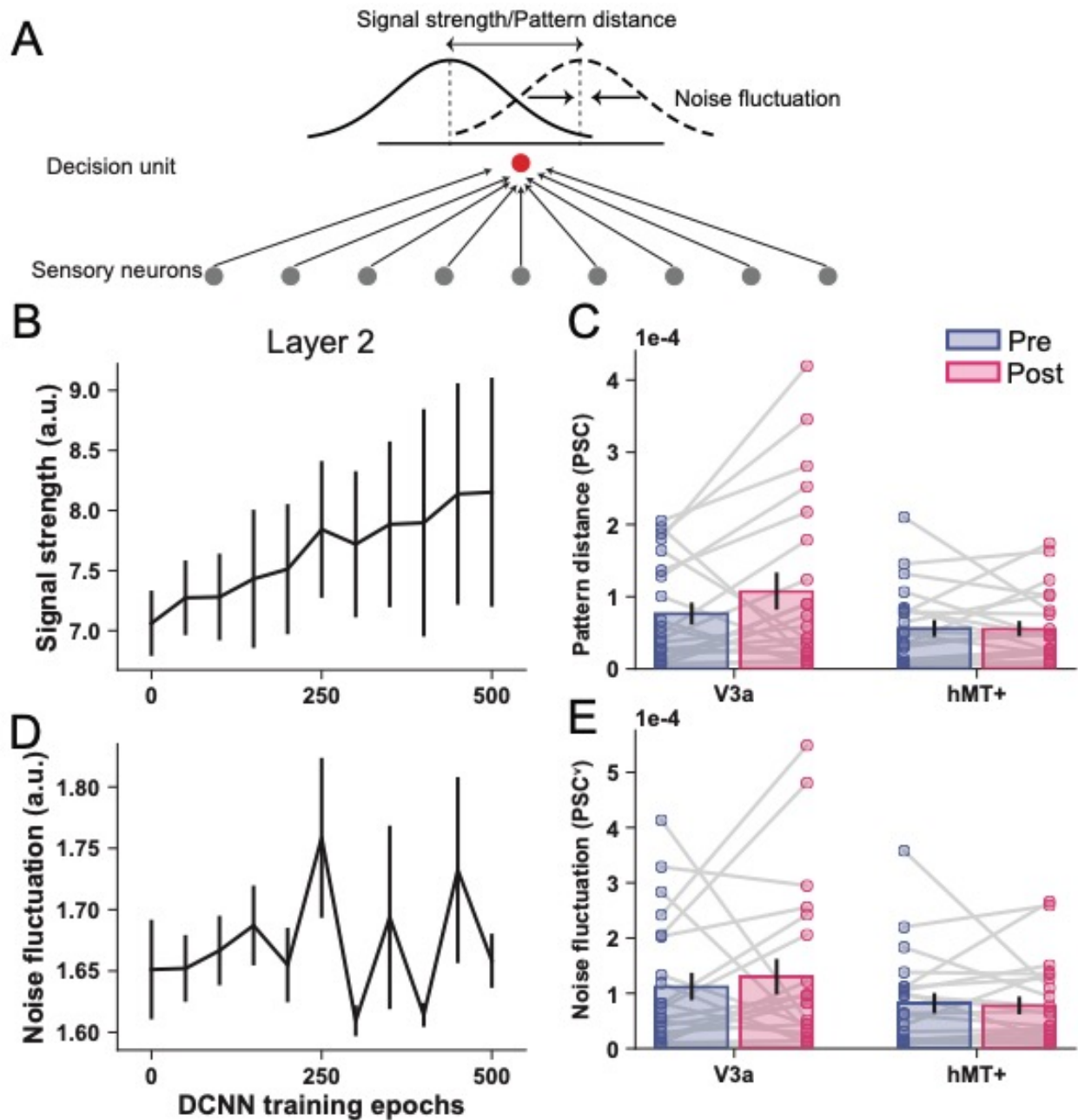


Figure S5. The effects of mechanism order on the results of stepwise information analyses. **A&B.** The cartoon illustration of the four mechanisms (**A**) and the results of the stepwise information analysis according to the order of SE/SE+MS/SE+MS+SR/SE+MS+SR+MW, the same as in Fig. 4 in the main text.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



99

Figure S6. We applied similar analyses in refs. [11](#)–[12](#) to our data. Here, we project the multivariate population responses onto an optimal decision unit and derive the two 1-d activity distributions of the decision unit for the two stimuli. We replicate the results that, after this linear transformation, the effects of VPL appear to be signal enhancement in layer 1 of the orientation DCNN (**A**), and in V3A but not in hMT+ in the human brain (**B**). In contrast, noise fluctuation in layer 1 (**C**) and in V3A and hMT+ in the human brain (**D**) do not appear to increase. These results are consistent with the Fig. 6C in ref. [11](#) and Fig. 6C in ref. [12](#). Data are presented as mean values \pm S.E.M., 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

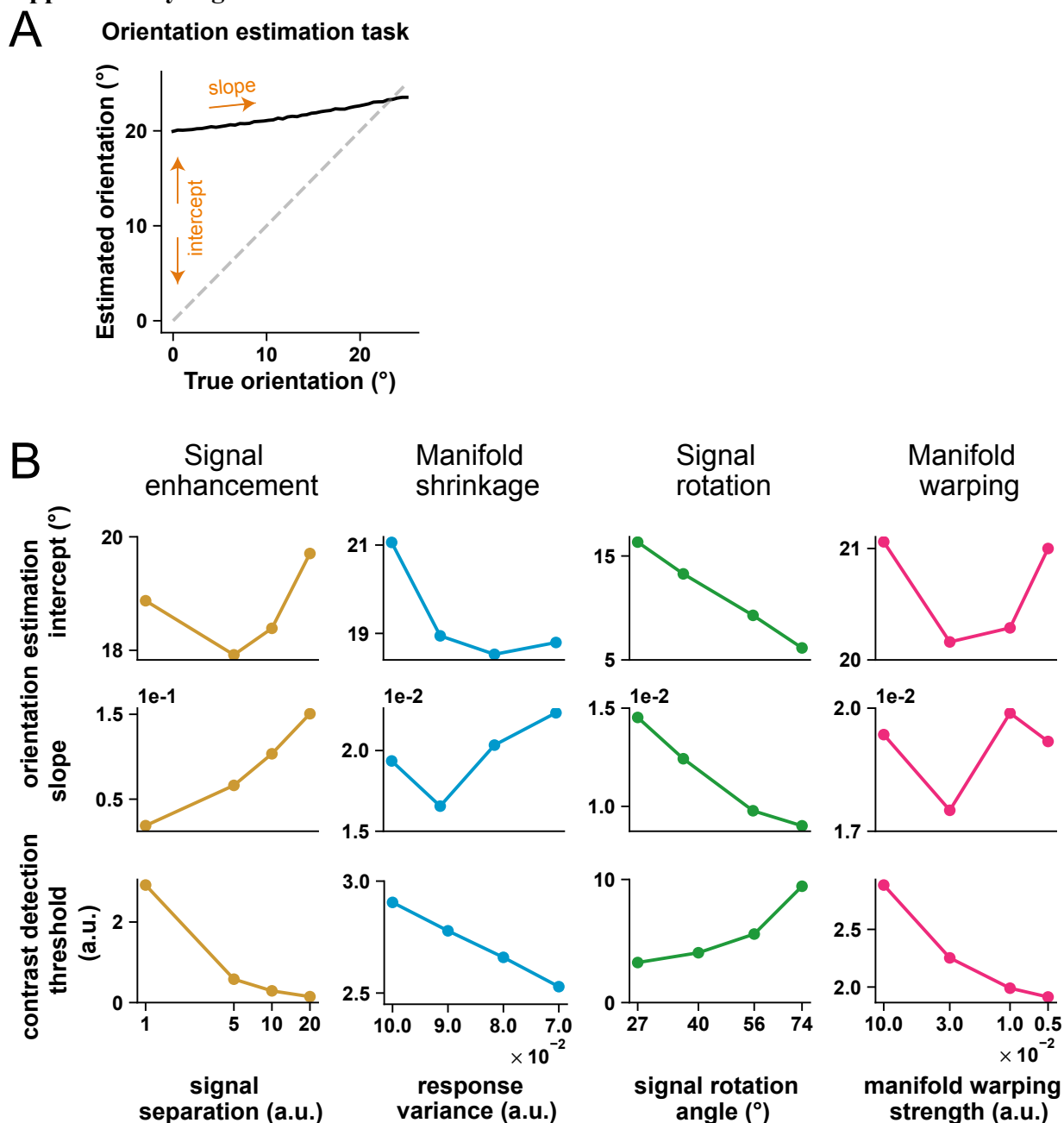


Figure S7. Predictions of the neural geometry approach on the effects of VPL of orientation discrimination on perceptual orientation bias reflected in orientation estimation and contrast detection. **A.** The estimated orientation as a function of the true presented orientation (solid line) in an orientation 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

Supplementary Table S1. Full statistical results of the **Figure. 2F** in the main text. All statistical 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	$[-\infty, -8.3 \times 10^{-2}]$

Supplementary Table S2. Full statistical results of the **Figure. 2G** in the main text. All statistical 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	$[-\infty, -7.7 \times 10^{-1}]$

Supplementary Table S3. Full statistical results of the **Figure. 3F** in the main text. All statistical 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	$[-\infty, 1.04]$

Supplementary Table S4. Full statistical results of the **Figure. 3H** in the main text. All statistical 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	$[1.5 \times 10^{-2}, \infty]$

Supplementary Table S5. Full statistical results of the **Figure. 4B** in the main text. All statistical 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	$[1.5, \infty]$

Supplementary Table S6. Full statistical results of the **Figure. 5C** in the main text. All statistical 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	$[-\infty, -7.4 \times 10^{-2}]$

Supplementary Table S7. Full statistical results of the **Figure. 5D** in the main text. All statistical 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	$[-\infty, -2.7]$

Supplementary Note 1: Additional explanations on the stepwise analyses

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

where A is the maximum response of this neuron, σ is the bandwidth of the tuning curve. μ_i is uniformly sampled from -90° to 90° . A total of 360 neurons are simulated. Additionally, a correlated gaussian noise with zero mean was added to each neuron.

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

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

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

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

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

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

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

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

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

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

Supplementary Note 4: Supplementary methods for the human fMRI experiment

Experimental stimuli and procedures. 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. Stimuli were generated and displayed using Psychtoolbox 3.0 in the MATLAB environment. To ensure consistency, the viewing distance in the behavioral sessions was set at 75 cm from the screen. The method for generating the random dot stimuli followed the approach described in ref. [18](#). All stimuli were presented within an imperceptible aperture of 10° diameter centered on a black background. At any given moment, the aperture contained 400 dots, each moving in unison in the same direction at a speed of 4 %/s. Dots that moved out of the aperture were repositioned on the opposite side of the aperture to maintain a constant dot density.

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

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

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

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

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

Supplementary Note 5: Supplementary methods for the macaque multielectrode recording experiment

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

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

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

For surgical preparation, animals were sedated with ketamine. During surgery, anesthesia and analgesia were maintained by sevoflurane (gaseous, 1–3%) and alfentanil (intravenous 156 µg/kg/h), respectively. Blood pressure, rectal temperature, blood oxygen saturation and end tidal CO₂ were measured continuously. After the surgery, analgesic (metacam 0.1/kg) and prophylactic antibiotics (cephorex 0.5 ml/kg) was given for 3–5 days.

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

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

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

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

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

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

Each trial was initiated when the monkey held a touch bar and fixated on a small fixation spot (diameter = 0.1° , fixation window = $2^\circ \times 2^\circ$) which was presented on a grey background (52.17 cd/m^2). During the trial, if the monkey broke fixation before saccade cue onset or failed to respond within 1000 ms of the onset of the saccade cue, the trial was terminated immediately and 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*

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

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

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

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

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

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

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

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

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

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

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

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

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

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 USA*. **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.* **4**e08998 (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