Running title: perceptual learning in biological and artificial visual systems

Noise reduction as a unified mechanism of perceptual learning in humans, macaques, and convolutional neural networks

Yu-Ang Cheng[1,2], Mehdi Sanayei[3,†], Xing Chen[4], Ke Jia[5,6,7], Sheng Li[8,9,10,11], Fang Fang[8,9,10,11], Takeo Watanabe[2], Alexander Thiele[3], Ru-Yuan Zhang[1]

[1]Institute of Psychology and Behavioral Science and Shanghai Mental Health Center, Shanghai Jiao Tong University, Shanghai, 200030, China.
[2]Department of Cognitive, Linguistic and Psychological Sciences, Brown University, RI, USA.
[3]Biosciences Institute, Newcastle University, Framlington Place, Newcastle upon Tyne NE2 4HH, UK.
[4]Department of Ophthalmology, University of Pittsburgh, 1622 Locust St, Pittsburgh 15239, USA.
[5]Department of Neurobiology, Affiliated Mental Health Center & Hangzhou Seventh People,
[6]Liangzhu Laboratory, MOE Frontier Science Center for Brain Science and Brain-machine Integration, State Key Laboratory of Brain-Machine Intelligence, Zhejiang University, Hangzhou, China.
[7]NHC and CAMS Key Laboratory of Medical Neurobiology, Zhejiang University, Hangzhou, China.
[8]School of Psychological and Cognitive Sciences, Peking University, Beijing, 100871, China.
[9]Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, 100871, China.
[10]PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing, 100871, China.
[11]Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing, 100871, China.
[†]Present address: School of Cognitive Sciences, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran

Ru-Yuan Zhang https://orcid.org/0000-0002-0654-715X

Corresponding:
Ru-Yuan Zhang
Zhongyuan 112, Shanghai Jiao Tong University
1954 Huashan Road, Xuhui District
Shanghai, China. 200030
Phone: +86-15674496106
Email: ruyuanzhang@stju.edu.cn

ABSTRACT

It is established that perceptual learning enhances the signal-to-noise ratios of sensory processing. However, the specific mechanisms behind this improvement remain enigmatic. Here, we systematically investigated these mechanisms by analyzing population activity changes in deep convolutional neural networks (DCNNs), humans, and macaques. We first trained DCNNs on orientation and motion direction discrimination tasks. Our models successfully replicated the behavioral signature of enhanced signal-to-noise ratio, and the individual and population response changes observed in numerous empirical studies. We then devised a method to quantify all scenarios in which learning modifies the tuning and covariance structure of neuronal populations to enhance sensory discrimination. A groundbreaking revelation is that enhanced signal-to-noise ratios are primarily driven by reduced trial-by-trial response variability. This finding was strongly supported by population response changes measured in visual areas of both humans and macaques. These results elucidate the shared mechanisms of neural plasticity in biological and artificial visual systems.

Keywords: visual perceptual learning, population codes, deep convolutional neural networks, signal-to-noise ratio, functional magnetic resonance imaging, multi-electrode recording

INTRODUCTION

Adaptively enhancing visual functions according to task demands is a critical ability of sensory organisms for maintaining longevity. It is well established in neuroscience that training on a visual task can produce long-lasting performance improvements, a phenomenon called visual perceptual learning (VPL) 1,2. Understanding how a perceptual system learns to perform new visual tasks is also a central goal of modern machine learning research. Although modern deep neural networks (DNNs) have achieved impressive performance in a variety of visual tasks, it remains unclear how learning refines the representations in DNNs and whether common mechanisms underlie perceptual learning processes in both biological and artificial visual systems.

Previous neuroscientific research on VPL generally falls into two broad categories. On the one hand, a number of theoretical studies have hypothesized enhanced signal-to-noise ratio as a powerful computational mechanism of VPL 3-5 (Fig. 1E). On the other hand, empirical studies have identified an array of neural phenomena associated with VPL, including sharpened orientation tuning curves 6,7, reduced firing rate variabilities (Fano factors) of individual V1 neurons and trial-by-trial co-variation of neuronal firing rate (noise correlations) in macaques 8-11, and increased sensory decoding accuracy in several human brain regions 12-14. Despite these tremendous advances, the connections between these two levels of research have not been thoroughly delineated. It remains unclear to what extent enhanced signal-to-noise ratios, as a algorithmic level theory of VPL, are implemented at the neural level, or, whether existing neural findings (e.g., sharpened tuning curves) are actually associated with enhanced signal-to-noise ratio in circuit computation.

Characterizing the neurocomputational mechanisms of VPL requires detailed measurements of the changes associated with learning in both single neurons and neuronal populations. A common approach is to record and contrast the activity of single neurons before and after learning. This approach provides the majority of the existing neurophysiological evidence, such as sharpened orientation tuning curves. However, it is controversial whether changes in single neuron responses actually improve stimulus encoding at the population level because there exist complex interactions between different aspects of responses (e.g., tuning curves, noise correlations) across different neurons on neural codes. For example, only a certain type of noise correlation is known to limit the information in a neural population 15. In addition, assessing the fidelity of sensory representations in population responses is notoriously difficult in both machine learning

and neuroscience. Although modern DNN models are capable of classifying nearly 1000 object categories (i.e., ImageNet), understanding the codes of these object categories in a large number of artificial neurons is hampered by complex nonlinear transformations. Developing statistical methods to interpret internal representations at the population level is an active area of research in machine learning 16. Despite the lack of statistical tools to understand population neural codes, an even greater challenge in the primate visual system is to simultaneously record the activity of the same population of neurons throughout the course of learning. Few studies have performed multi-unit recordings during perceptual learning 8,10,11. Direct empirical evidence for how single-neuron responses improve population codes is still lacking.

The above challenges suggest that, to better understand the neurocomputational mechanisms of VPL, we need a unifying approach that achieves at least two objectives: (1) specify how theories at the algorithmic level, such as enhanced signal-to-noise ratio, manifest in population responses, and (2) provide a quantitative account for linking the effects of individual and population responses on neural codes in the context of perceptual learning.

Here, we introduce a comprehensive computational framework that decomposes improved population codes for sensory discrimination into four possible mechanisms. Two of them (i.e., signal enhancement and noise reduction) are the direct extension of the one-dimensional signal detection theory. Two other mechanisms (i.e., signal rotation and covariance warping) can only be revealed in a high-dimensional neural space and they specify how learning alters the representational geometries of stimuli. To test this framework, we first train deep convolutional neural network (DCNN) models on two classical VPL tasks—orientation and motion direction discrimination. The DCNN models reproduce a wide range of empirical psychophysical, human neuroimaging, and macaque neurophysiological findings. We apply our analytical framework to the population responses in the DCNN models. Interestingly, the results show the existence of three mechanisms proposed by our framework, and implicate noise reduction as the primary mechanism of VPL. These findings in the DCNN models are further tested on cortical response changes associated with motion direction discrimination learning in humans and electrophysiological population response changes associated with contrast discrimination learning in macaque V4. Our study presents a general approach to understanding neural codes in single neurons and neural populations, and how cognitive processes (e.g., learning) affect neural codes to produce more efficient sensory representations.

RESULTS

Orientation discrimination learning improves behavioral performance of DCNN

To elucidate the neurocomputational mechanisms of VPL, we trained a DCNN model (Fig. 1A) to perform a classical orientation discrimination task 4. Similar to the model in ref. 17, this model inherits the first five convolutional layers of AlexNet, which was pre-trained on ImageNet 18. 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 (i.e., clockwise or counterclockwise rotation of the target stimulus relative to the reference stimulus). Importantly, similar to previous psychophysical studies 4,19, we systematically manipulated the level of input image noise (Fig. 1B). The model was trained on stimuli with multiple noise and contrast levels (see Methods and Materials for training details).

To evaluate model performance and compare it directly with human psychophysical results, we assessed orientation discrimination accuracy as a function of stimulus contrast and noise (i.e., behavioral psychometric functions in Fig. 1C&D), and further derived contrast thresholds as a function of image noise level (Fig. 1F, Threshold vs. Noise, TvN function). We found that training significantly improved model performance in this task in almost all stimulus contrast and noise conditions. Importantly, training reduced contrast thresholds at all noise levels (Fig. 1F). This downshift of TvN functions is consistent with the well-established psychophysical result (replotted in Fig. 1E) as previously observed in humans 4,5. This signature change of TvN functions has also been treated as a hallmark of enhanced signal-to-noise ratio in sensory processing induced by perceptual training.

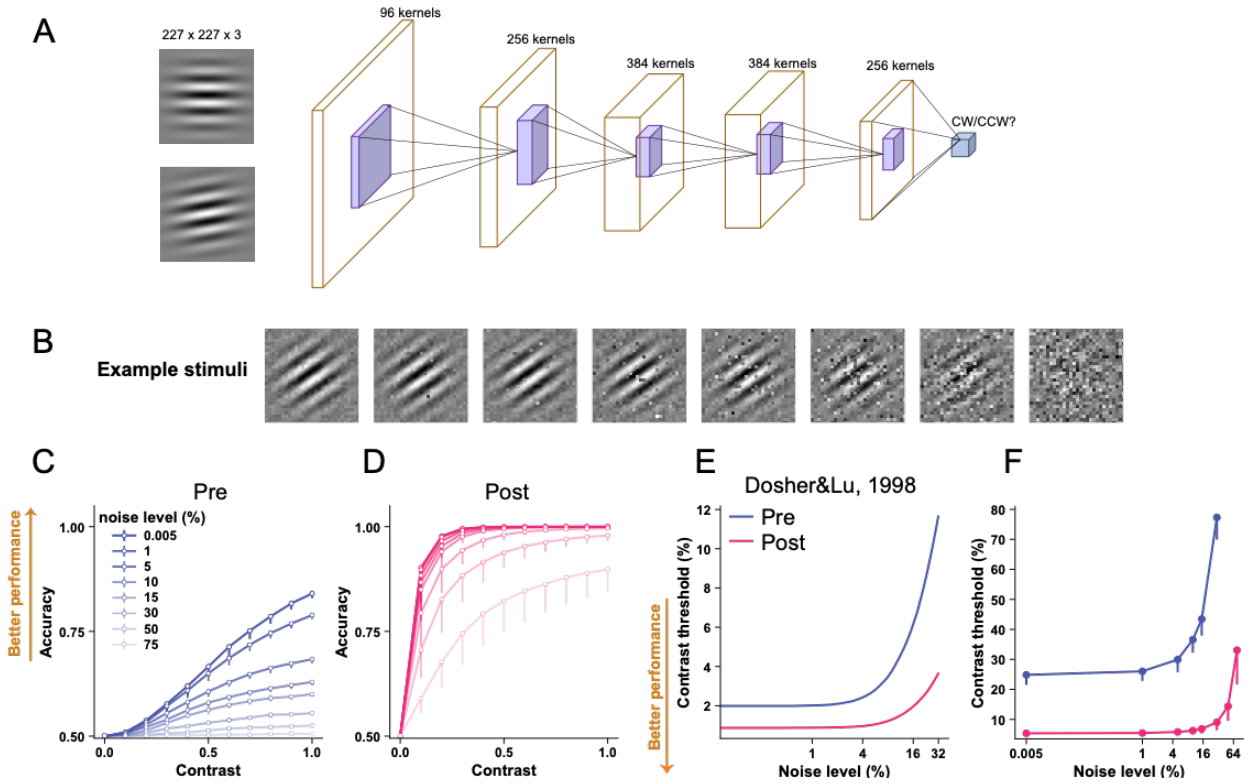

Figure 1. A DCNN model (***A***) is trained on an orientation discrimination task with Gabor stimuli embedded in different levels of image noise (***B***). Orientation discrimination accuracy is significantly improved from pre- (***C***) to post-test (***D***). Training induces a downshift of the Threshold vs. 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. 4). The absolute quantitative differences between ***E*** and ***F*** may be due to differences in the overall signal-to-noise ratio or the number of layers and neurons between the human visual system and the DCNN model.

Orientation discrimination learning refines sensory representations in neural populations

Having established the behavioral signature of perceptual learning in our DCNN model, we next sought to understand the effects of VPL on population coding in the model. We simulated many trials elicited by the two orientation stimuli for fine discrimination and performed multivariate decoding analyses in each layer. Model training improved decoding accuracy in later layers (layers 3-5, Fig. 2F). The learning effect is more pronounced in later layers possibly because the benefits of VPL accumulate from early to later layers. More formally, we calculated linear Fisher information, a classical metric in computational neuroscience, to quantify stimulus information (i.e., how well the two stimuli can be discriminated based on population responses, see Methods and Materials) in each layer before and after VPL. The amount of sensory information represented

in later layers was indeed enhanced by VPL (Fig. 2G). Such refined sensory representation at the population level is consistent with the decoding results based on both macroscopic cortical activity in humans [12,13,20] and multiunit spiking activity in macaques [10,11].

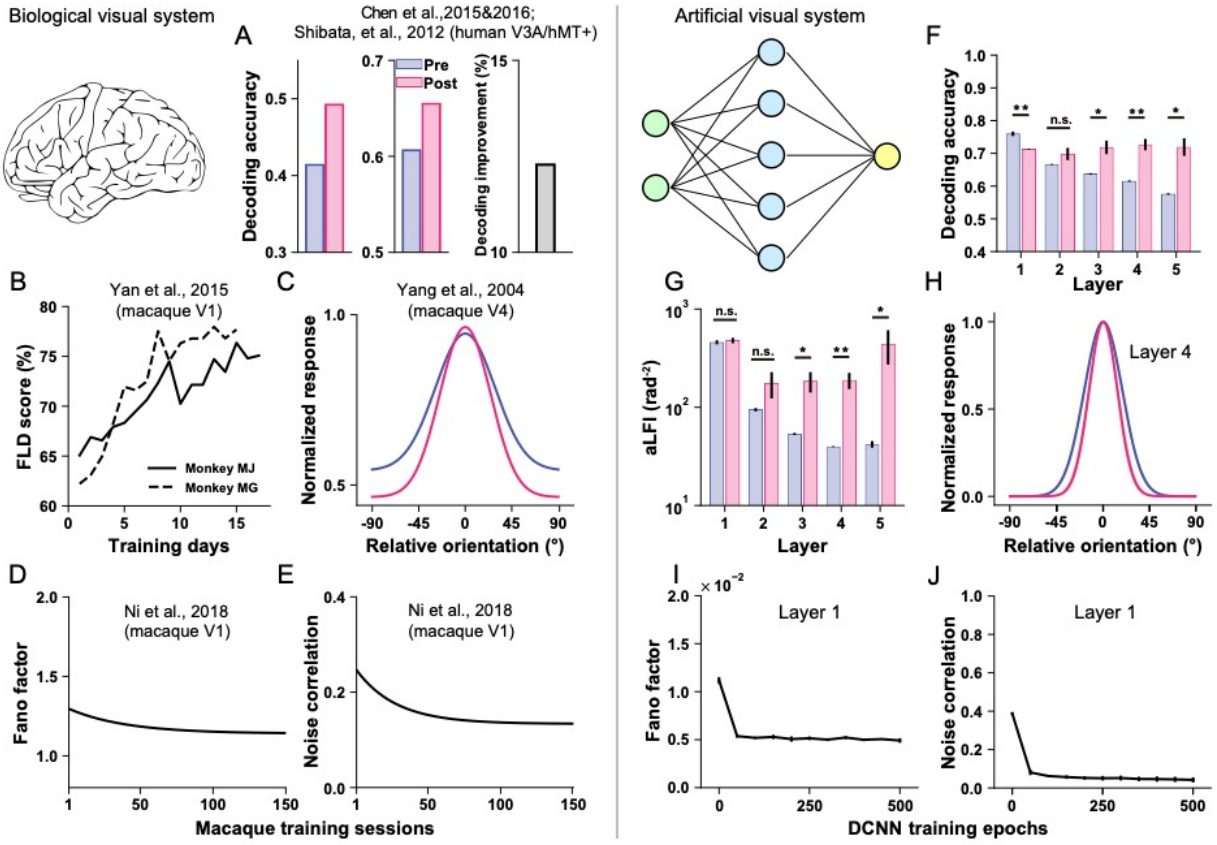

Figure 2. Neural consequences of VPL in humans (***A***), macaques (***B-E***), and in our DCNN model (***F-J***). VPL improves stimulus decoding accuracy in related regions in the human brain (***A***) and decoding scores of Fisher's linear discriminant (FLD) in macaque V1 (***B***). VPL sharpens orientation tuning curves of neurons in macaque V4 (***C***), and also reduces Fano factors and interneuron noise correlations (***D&E***). Similar results are observed in the DCNN model: network training also improves decoding accuracy in layers 3-5 (***F***), and averaged linear Fisher information (aLFI, total information in each layer divided by the number of neurons in that layer) (***G***) in layers 3-5. VPL sharpens orientation tuning curves of neurons in layers 1-4 in the DCNN model (results of layer 4 only are shown in ***H***). Similar reduction of Fano factors and noise correlations are observed in the DCNN model (results of layer 1 only are shown in ***I&J***). The data shown in ***H-J*** are the median value across neurons in a layer. The results of all five layers are shown in Supplementary Fig. S1. Panels ***A-E*** are reproduced by the data points shown in the original papers. Error bars and error shadings in panels ***F-J*** represent the S.E.M. across four reference orientations (error shadings in ***H*** are small and barely visible).

185 ***Orientation discrimination learning alters response properties of individual neurons***

186 In addition to characterizing VPL on the population level, we sought to investigate the neural
187 consequences at the single-unit level.

188 Importantly, we found that three key neural signatures of VPL as documented in the
189 neurophysiological literature emerge naturally from the neural network training. First, training
190 modestly sharpened the tuning curves of artificial neurons in layers 1 to 4 (Fig. 2H& and
191 Supplementary Fig. S1). The finding of altered tuning curves by perceptual training has been
192 reported previously in refs. 6,7,21 (Fig. 2C, but see also null results in ref. 22). Second, we observed
193 a significant decrease in Fano factor of individual neurons in all five layers (Fig. 2I and
194 Supplementary Fig. S1), a phenomenon indicating an increased signal-to-noise ratio of individual
195 neuronal responses. This has been reported in both humans 23 and macaques 8,21 (Fig. 2D). The
196 sharpened tuning curve and reduced Fano factor are also consistent with theoretical modeling 17.
197 Third, training reduced trial-by-trial noise correlations between neurons in all five layers (Fig. 2J
198 and Supplementary Fig. S1), a finding also consistent with several empirical results in macaques
199 8-11. Critically, we also found that the reduction in noise correlation depended on tuning similarity.
200 Learning reduced the noise correlations between neurons with similar tunings (i.e., positive signal
201 correlations) and increased the noise correlations between neurons with opposite tunings (i.e.,
202 negative signal correlations) (Supplementary Fig. S2). Previous theoretical work has suggested
203 that the former type of noise correlations is detrimental for information coding and the latter type
204 is beneficial 15,24. The pattern of reduced detrimental and increased beneficial noise correlations
205 has been discovered with learning tasks in songbirds 25 and with attention tasks in macaques 26.

206 In addition to these classical neurophysiological findings in VPL, our model also captures
207 some important response properties of sensory neurons in the primate early visual system. First,
208 the relationship between the Fano factor and orientation tuning of the artificial neurons bears
209 strong resemblances to empirical measures of V1 neurons in macaques 27 (Supplementary Fig.
210 S2). Second, we found a positive relationship between signal correlation and noise correlation
211 among artificial neurons in all layers (Supplementary Fig. S2). This relationship has also recently
212 been documented as a ubiquitous phenomenon in both electrophysiological 28-30 and human
213 imaging 31-33 studies.

214 Taken together, our neural network model reproduces a wide range of psychophysical,
215 human imaging, and macaque neurophysiological findings in VPL. Although these similarities are

qualitative and not quantitative (see Discussion), the results suggest that the DCNN model allows us to explore neurocomputational mechanisms that may be difficult to elucidate in empirical experiments.

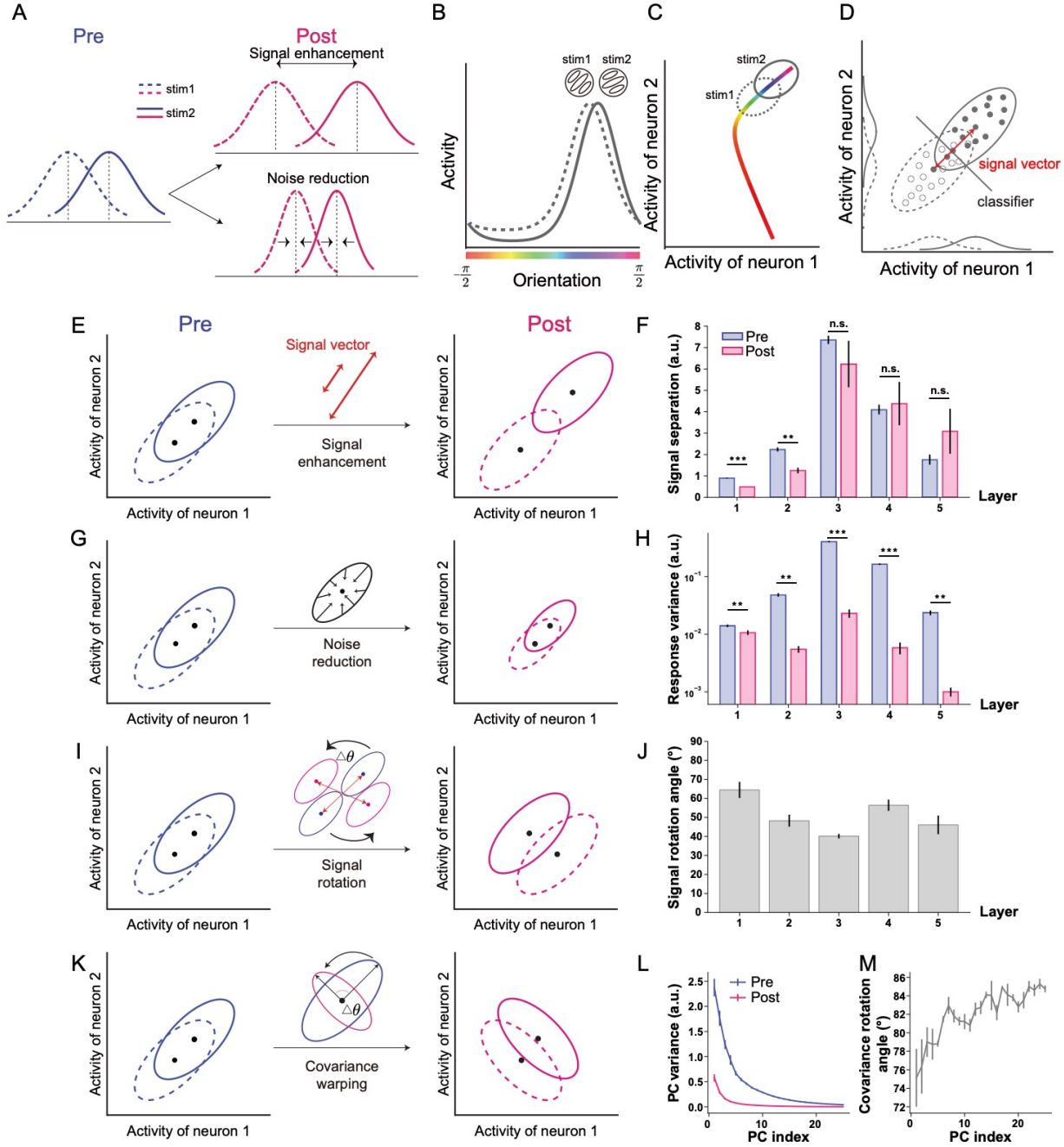

Figure 3. Four possible mechanisms of VPL in neural populations. According to classical signal detection theory, better sensory discriminability can be represented as a larger separation between two response distributions. Signal enhancement predicts enlarged distances between the mean while noise reduction predicts reduced variance between the two response distributions (***A***). Stimulus orientation as a continuous stimulus variable can evoke population responses with the same shape as individual tuning curves (***B***). If we

continuously sweep the orientation value, the mean of the population responses forms a manifold in a high-dimensional neural space with the number of dimensions being equal to the number of neurons. The mean population response to the two stimuli in a discrimination task are two points on the manifold (***C***). In realistic population responses, the trial-by-trial population responses to the two stimuli form two high-dimensional response distributions. The distributions look elliptical rather than spherical due to pairwise noise correlations between neurons (***D***). The signal enhancement mechanism predicts an increased Euclidean distance (i.e., signal separation, ***E***) between two high-dimensional response distributions. However, no significant increase in signal separation is observed in any of the five layers (signal separation decreases in the first two layers, ***F***). The noise reduction mechanism predicts reduced variance (***G***) of the two high-dimensional response distributions. This is observed in all five layers (***H***). The signal rotation mechanism predicts that the positions of the mean of the two distributions are changed by VPL (***I***). The rotation angle ranges from 50º~70º in all five layers (***J***). The covariance warping mechanism predicts that VPL changes the shape of the population response covariance (***K***). Indeed, VPL reduces the variance of the high-variance principal components of the population responses (***L***). The principal components (PC, only show PCs that account for >99% of the total variance) are ranked from high to low variance. More importantly, the directions of the principal components rotate significantly from pre- to post-test (***M***).

Neural correlates of four computational mechanisms of VPL in population responses

How would enhanced signal-to-noise ratio in sensory processing manifest in high-dimensional population responses? In the simplified one-dimensional scenario (Fig. 3A), classical signal detection theory posits that better sensory discrimination can be achieved by either increasing the distance between the means (i.e., signal enhancement) and/or decreasing their variance (i.e., noise reduction), because the former only alters the tuning (i.e., mean) of a response unit and the latter only reduces the variance around a fixed mean. In multivariate population responses, the two stimuli to be discriminated instead generate two multivariate response distributions in a high-dimensional neural space whose dimension corresponds to the dimension of a population (Fig. 3B&C). In a simplified visualization in 2D space (i.e., assuming a population of only two neurons, Fig. 3D), the two distributions are elliptical due to noise correlations between neurons. We refer to the vector connecting the mean of the two distributions as the *signal vector* and its modulus length (i.e., the Euclidean distance between the two stimulus distributions) as the *signal separation*.

In the high-dimensional neural space, we illustrate four possible mechanisms by which VPL may improve the discrimination of two stimuli. First, according to the classical signal detection theory, the *signal enhancement* mechanism predicts an increased Euclidean distance between the two response distributions (Fig. 3E). However, we found that the signal separation

between the two distributions did not significantly increase with learning in all five layers, and even slightly decreased in the first two layers (Fig. 3F). Second, consistent with the classical signal detection theory, the *noise reduction* mechanism predicts that VPL reduces the trial-by-trial response variance of neurons (Fig. 3G). This is what we found in all five layers (Fig. 3H). These results support the hypothesis that noise reduction plays a more important role than signal enhancement in VPL. We further included two novel mechanisms that can only occur in high-dimensional neural space and lead to increased discriminability of two response distributions. In the third mechanism, although VPL did not increase signal separation, it may change the positions of the mean of the two distributions and consequently increase discriminability due to the elliptical shape of the distributions (Fig. 3I). Interestingly, we found that the signal vectors in each layer were rotated by ~50°-70° after training (Fig. 3J). We call this mechanism *signal rotation*. Finally, covariance structures play a pivotal role in high-dimensional response distributions. We found that VPL systematically warped the shape of the covariance structures of high-dimensional response distributions (Fig. 3K-M). We refer to this mechanism as *covariance warping*. Note that covariance warping includes both the changes in correlation structures and the redistribution of variances across individual neurons, while holding the total variance constant. It is noise reduction that attenuates the total variance.

In summary, here we present four possible neural mechanisms of VPL and identify neural evidence for three of them (i.e., except signal enhancement) in our DCNN training. Note that these four mechanisms (or a combination thereof) encompass all possible scenarios by which learning can potentially improve sensory discrimination.

Information-theoretic analyses reveal the key role of noise reduction in orientation discrimination learning

Given the four possible mechanisms (i.e., signal enhancement, noise reduction, signal rotation, and covariance warping) and their complex interaction effects, how can we delineate their respective contributions to improved population codes? Here, we introduce a stepwise approach based on the notion of linear Fisher information. 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. 4A). For example, as shown in Fig. 4, we first calculate information gain by considering only the signal enhancement scenario, then by

considering both signal enhancement and noise reduction, and so on until all four mechanisms are included. We emphasize that this stepwise approach is intended only to facilitate the understanding of the computational underpinning of each mechanism. It does not imply that these mechanisms must occur sequentially over the course of VPL.

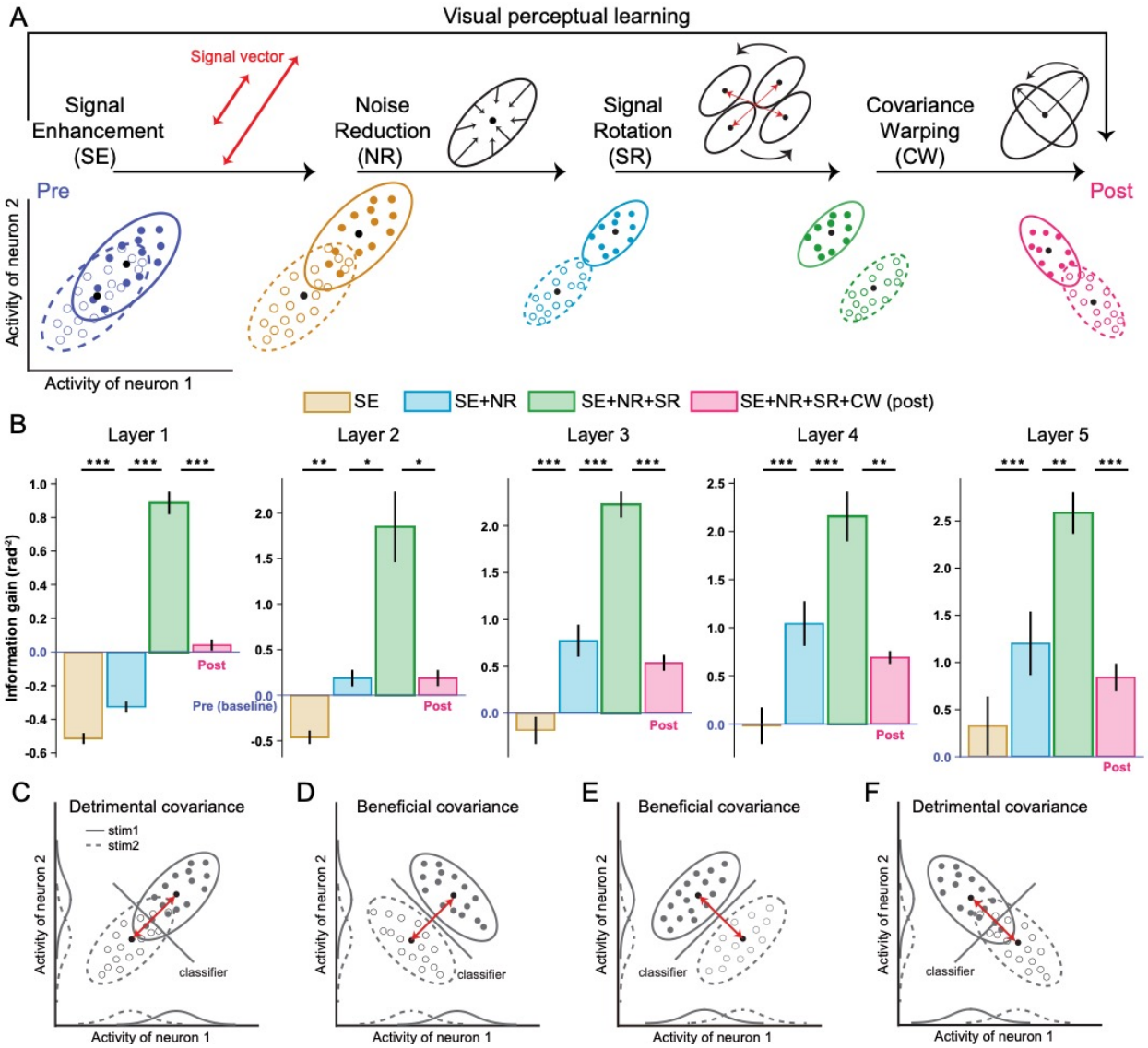

Figure 4. The effects of four mechanisms on population codes are decomposed into four distinct steps (***A***). ***B*** shows the effects on information gain by sequentially adding each of the four mechanisms in each layer. For example, the increase in height from the brown to the blue bars indicates the positive contribution of noise reduction to encoded stimulus information. ***C-F*** illustrate the 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 the signal vector. Similarly, the effects of the signal vector also depend on its relative geometry to the axis of covariance.

Interestingly, not all mechanisms have a positive effect on population codes. We found that the effect of signal enhancement is minimal in all five layers. This mechanism even reduces stimulus information in layers 1&2. This is consistent with the reduced Euclidean distance in the first two layers (Fig. 3F). Noise reduction significantly enhances stimulus information in almost all layers. This makes sense because the decreased variance reduces the overlap between the two response distributions. 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 covariance warping is further considered (Fig. 4B, magenta bars) because VPL also warps the covariance direction to realign it with the post-test signal vector, thereby reducing stimulus information.

We want to highlight two important aspects of such stepwise analyses. First, we use signal rotation and covariance warping to reflect changes in the direction of the signal vector and the shape of the covariance, the two types of changes that can theoretically occur without each other (Fig. 3I&K). It has been shown that the relative geometric relationship between covariance and signal vector rather than their absolute magnitude or shape determines sensory discriminability [15,24] (Fig. 4C-F). The fact that learning changes both signal vector and covariance further complicates the matter. We therefore argue that their contributions to improved population codes should be evaluated together. For example, in Fig. 4B, signal rotation and covariance 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, it can be argued that noise reduction is the primary contributor to improved stimulus information.

Second, one might ask whether the results might manifest differently if the order in which the mechanisms are applied changes. As mentioned above, the effects of signal rotation and covariance warping should be evaluated together, but the effects of signal enhancement and noise reduction are independent. Hence, changing the order does not affect our results as long as signal rotation and covariance warping are carried out consecutively. In fact, the apparent positive and negative effects of signal rotation and covariance warping in the stepwise analyses can be reversed when their order is switched, but both orders actually reflect identical underlying changes (see detailed analyses in Supplementary Fig. S6).

Taken together, our neural network model allows us to readily measure the activity of all our simulated neurons across the visual hierarchy and to perform systematic information-theoretic analyses to disentangle the effects of different neural consequences after VPL. Importantly, we found that reduced noise in population responses was the key mechanism underlying the improved population codes induced by VPL in the DCNN.

Noise reduction as the primary mechanism of motion direction discrimination learning in DCNN

The above analyses focus only on a 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 34,35, human imaging 12,13, and neurophysiological studies 36. Importantly, motion perceptual learning 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 pre-trained C3D network 37 and trained the model to perform a motion direction discrimination task commonly used in psychophysics (see Methods and Materials for stimulus and training details).

In the motion DCNN model, we found similar mechanisms as in the orientation discrimination task. First, training significantly improved the behavioral performance of the model (Fig. 5B). Second, training also enhanced decoding accuracy and averaged linear Fisher information in later layers (Fig. 5C&D), suggesting that motion VPL refines stimulus representation at the population level. Third, the effects of motion VPL on individual neurons in layer 6 are also pronounced (see results for all six layers in Supplementary Fig. S3). We found that motion VPL reduced Fano factor (Fig. 5E) and noise correlations (Fig. 5F). Fourth, motion VPL did not significantly improve signal separation (Fig. 5G) but markedly reduced response variance (Fig. 5H). In addition, motion VPL also induced two novel mechanisms—signal rotation (Fig. 5I) and covariance warping (Fig. 5J&K). Most importantly, the four mechanisms induced by motion VPL had similar respective contributions to population codes (Fig. 5L).

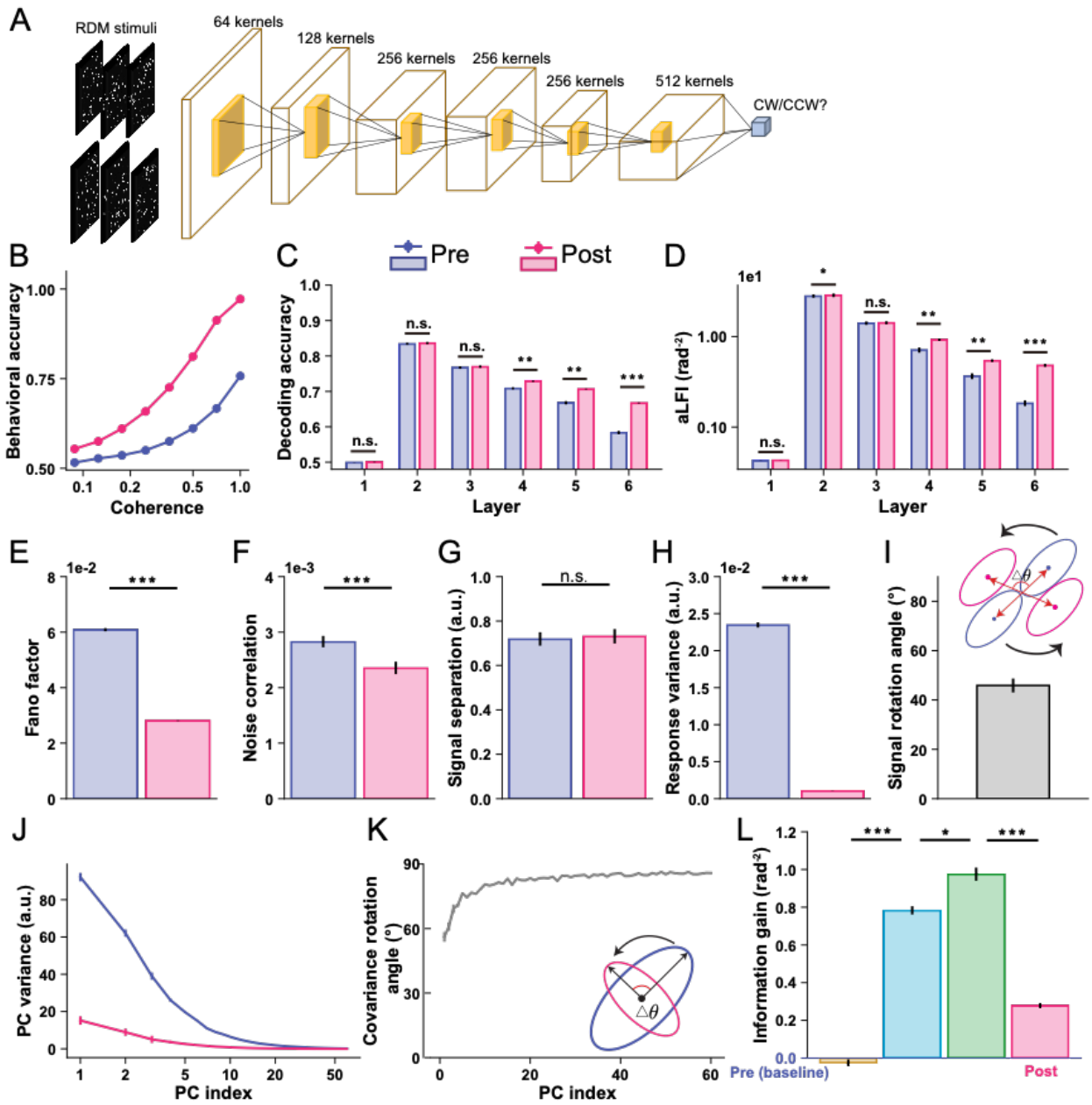

Figure 5. Results of DCNN modeling of motion VPL. The DCNN model (***A***) of motion VPL uses 3D convolutions to process video stimuli. Here, we simplify the 4D feature maps in each convolutional layer and show them as 3D maps only for illustration purposes. Motion VPL significantly improves DCNN direction discrimination performance (***B***), decoding accuracy (***C***), and averaged linear Fisher information (***D***). For single neuron analyses, motion VPL also reduces the Fano factor (***E***) and noise correlation (***F***) in layer 6. Similar to orientation VPL, motion VPL does not significantly enhance signal separation (***G***), but significantly rotates the position of the two distributions in layer 6 (***I***). Importantly, VPL clearly reduces the response variance in layer 6 (***H***). Specifically, motion VPL reduces the variance of the high variance PCs and rotates the directions of all PCs, indicating a significant effect of covariance warping in layer 6 (***J***&***K***). The pattern of information gain associated with the four possible mechanisms (***L***) is consistent with that of orientation VPL.

See results for all six layers in Supplementary Fig. S3. Note that some error bars are very small and barely visible.

Noise reduction as the primary mechanism of motion direction discrimination learning in the human brain

The converging results in the DCNN models of orientation and motion direction discrimination, and the remarkable agreement between our DNN models and existing empirical neuroscientific findings support the biological plausibility of our models. Our DCNN framework additionally reveals some novel mechanisms (i.e., signal rotation and covariance warping) and suggests that noise reduction is the primary mechanism of VPL. However, it remains unknown whether these predictions are only present in the DCNN models and have no biological basis in the brain.

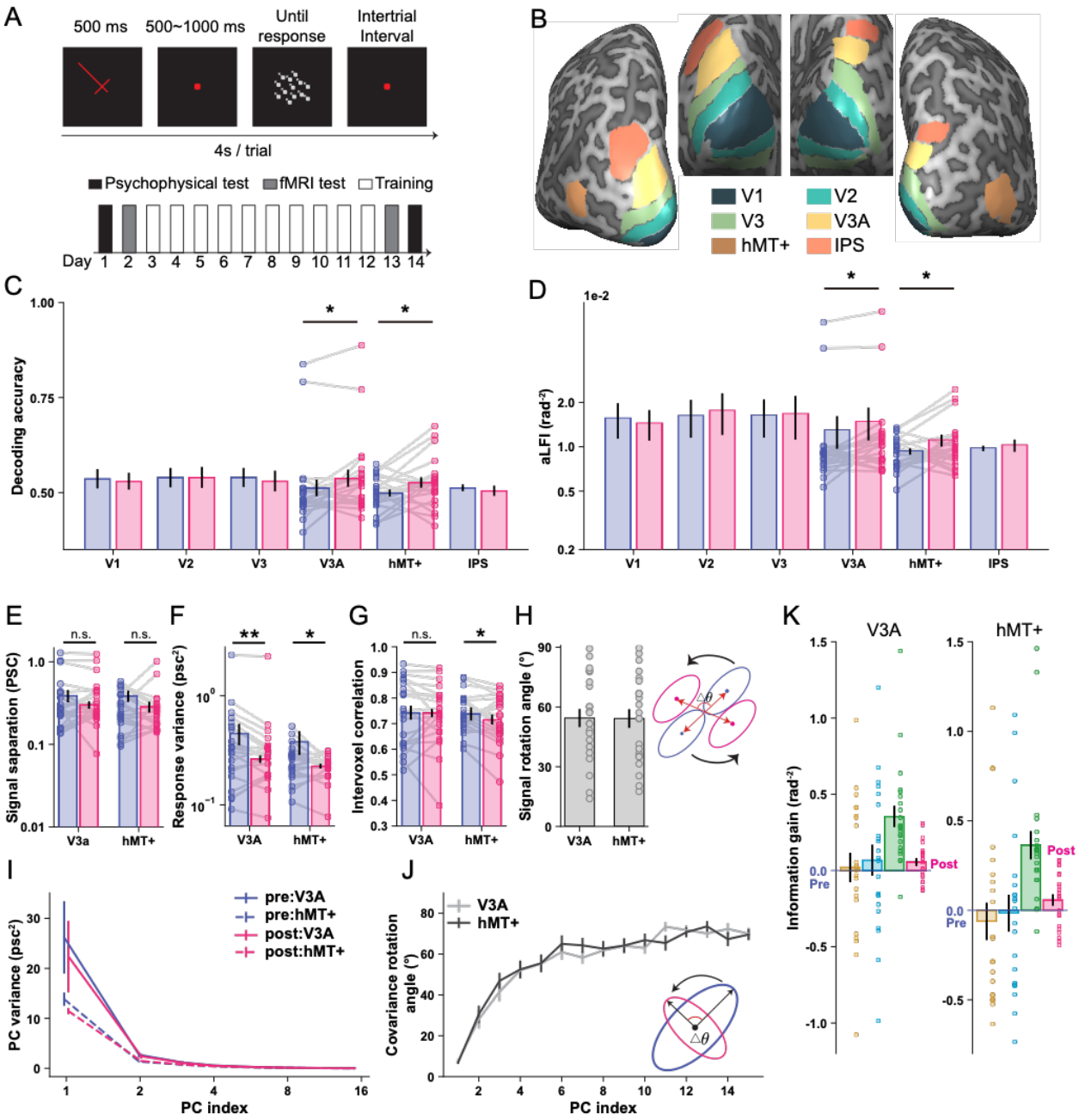

Figure 6. Trial diagram and training paradigm (***A***), and regions-of-interest in a typical subject (***B***). Motion VPL in humans significantly improves decoding accuracy (***C***) and averaged linear Fisher information (***D***) in areas V3A and hMT+, a finding consistent with several existing fMRI studies of motion VPL. Note that the four data points in V3A appear as outliers in ***C***&***D***, but the results still hold if these data points are removed. Motion VPL does not significantly alter signal separation in V3A and hMT+ (***E***), but reduces voxel response variance in V3A and hMT+ (***F***) and intervoxel noise correlations in hMT+ (***G***). Similar to the DCNN models of VPL, motion VPL in humans also rotates stimulus distributions (***H***), reduces the variance of high variance PCs (***I***), and warps the covariance directions (***J***). The patterns of information gain associated with the four mechanisms (***K***) are consistent with those in the DCNN

models of both orientation and motion VPL. The unit "PSC" represents percent signal change of blood oxygenation level dependent signals. Individual data points represent the human subjects. Error bars in all panels represent the S.E.M. across subjects. Significance conventions are: *, $p < 0.05$; **, $p < 0.01$.

To address this question, we analyzed blood oxygenation level dependent (BOLD) responses in the cortex of human subjects before and after they were trained on a motion direction discrimination task (Fig. 6A, ref. 38). 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 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, and found that perceptual training significantly enhanced decoding accuracy (Fig. 6C) and averaged linear Fisher information (Fig. 6D) in areas V3A and hMT+, a result consistent with several human fMRI studies on motion VPL 12,13,20.

We further investigated the coding principles in areas V3A and hMT+, and repeated the above analyses of DCNN models on fMRI data. Note that we performed the same analyses on voxels rather than artificial neurons in DCNN models. Consistent with the predictions of the DCNN models, motion VPL in humans did not increase signal separation (Fig. 6E) but markedly reduced voxel response variance (Fig. 6F) in both areas. Motion VPL also significantly reduced intervoxel correlations in hMT+ (Fig. 6G). The mechanism of signal rotation was also evident, as indicated by the average ~55º rotation of the signal vectors in both areas (Fig. 6H). In addition, motion VPL warped the magnitude and direction of the covariance (Fig. 6I&J). Most importantly, the respective contributions of these four mechanisms in both brain regions were similar to the pattern in the DCNN models (Fig. 6K).

Noise reduction as the primary mechanism of contrast discrimination learning in macaque V4

Voxel responses reflect macroscopic brain activity that aggregates the responses of ~300, 000 – 50, 000 neurons 39. It remains unclear whether the mechanisms we have discovered so far also exist at the local circuit level of single cells or small clusters of cells. Although our DCNN models

reproduce several neural phenomena in single neurons, such as reduction of noise correlations and Fano factors, they also predict that (1) VPL induces novel mechanisms (e.g., signal rotation and covariance warping) that can only be revealed by analyzing multivariate responses, and that (2) noise reduction is the primary mechanism of VPL. To our knowledge, these predictions have not been systematically tested using intracranial recording.

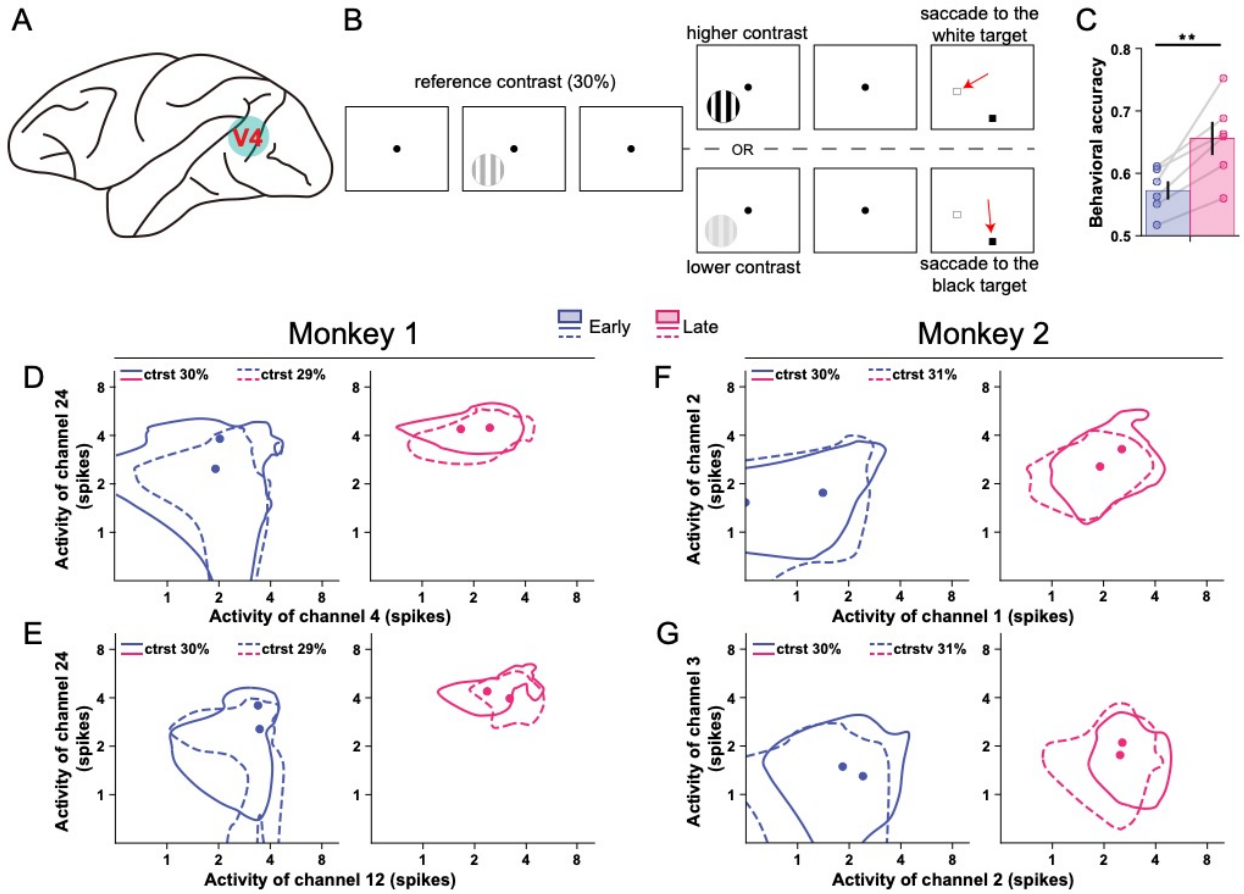

Figure 7. We analyzed population responses in area V4 (***A***) of two monkeys while they were trained on a fine contrast discrimination task (***B***). The first four and last four training sessions were grouped as pre- and the post-test conditions, respectively. VPL significantly improves behavioral performance from the early to late stage of training (***C***). All individual data points in ***C*** 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. S5. Error bars indicate the S.E.M. across the six conditions. ***D-G*** illustrates the full width at half maximum of the response distributions of four pairs of channels at pre- and post-test (***D&E*** for monkey 1 and ***F&G*** for monkey 2). Solid lines represent 30% reference contrast, and dashed lines represent 29% and 31% target contrast in monkey 1 and monkey 2, respectively. These results show that learning systematically warps the geometries of the multivariate responses.

To further test our hypotheses on neuronal spiking activity, we analyzed the population responses of V4 neurons in two monkeys (Fig. 7A) at the early and the stage of training in which they learned to perform on a fine contrast discrimination task (Fig. 7B, ref. 11). In this task, each monkey was presented sequentially with two identical Gabor patches with different contrast levels. The contrast of the reference (i.e., the first) stimulus was always fixed at 30%, and the contrast of the target (i.e., the second) stimuli varied systematically near the reference contrast (i.e., 27/28/29/31/32/33%). This contrast discrimination training significantly improved behavioral performance (Fig. 7C). Most importantly, responses of multiple channels were continuously recorded via chronically implanted electrodes in area V4 (21 and 29 channels for monkey 1 and 2, respectively) throughout training (21 and 23 training sessions for the two monkeys, respectively). This continuous multi-electrode recording is key to disentangling population-level changes in information encoding during VPL.

We used the above analyses (previously applied to DCNN models and human fMRI data) and applied them to the macaque V4 responses, and again found highly consistent results. First, contrast discrimination training improved stimulus information at the population level (Fig. 8A&B). Second, at the individual level, VPL also significantly reduced Fano factors (Fig. 8C) and noise correlations (Fig. 8D), consistent with several existing findings. Interestingly, while the trial-by-trial variance in firing rate was significantly reduced after training (Fig. 8F), no apparent change in signal separation was observed (Fig. 8E), suggesting the predominant role of noise reduction. Importantly, we again observed evidence for signal rotation (Fig. 8G) and covariance warping (Fig. 8H&I). The stepwise information analyses also qualitatively replicated the relative contributions of the four mechanisms to the total stimulus information encoded in the population (Fig. 8J).

Taken together, we found highly consistent results across species, behavioral tasks, experimental conditions, measurement modalities, and brain regions. These converging results in humans, macaques, and DCNN models suggest that both biological and artificial systems share common mechanisms of VPL, and that noise reduction is the primary contributor to VPL.

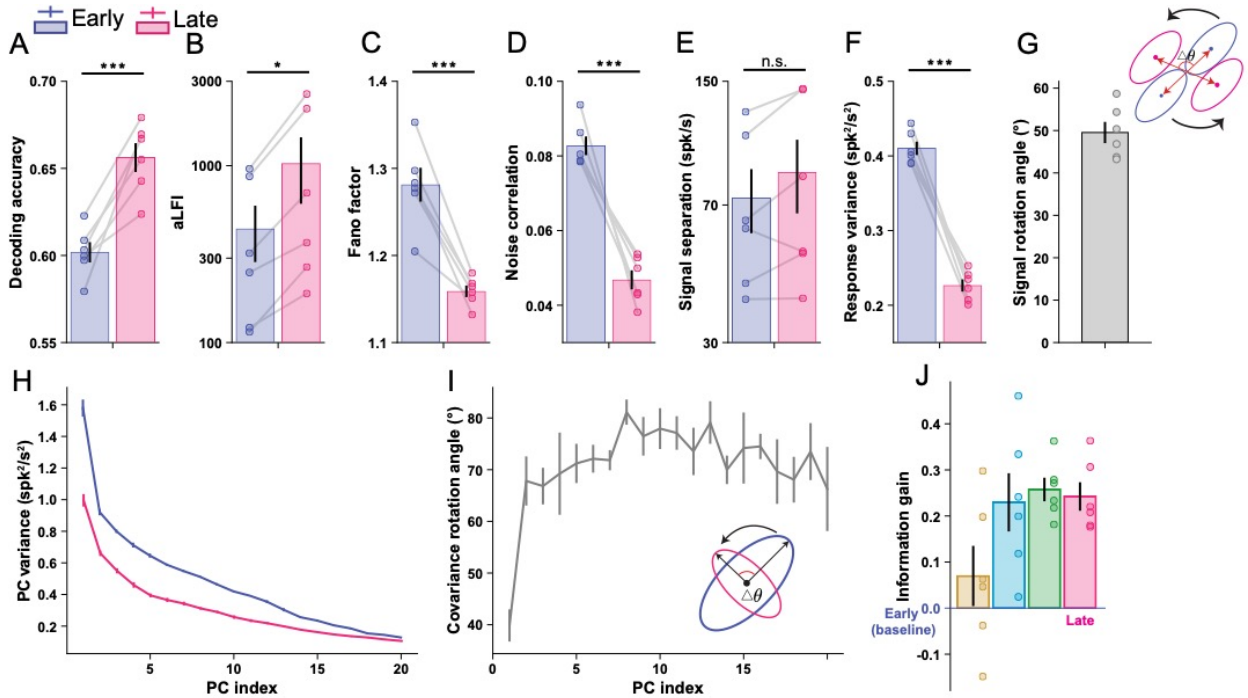

Figure 8. Population activity analyses of contrast discrimination learning in macaque V4. VPL significantly improves stimulus information at the population level (***A***&***B***). Consistent with VPL in the DCNN models and the human brain, VPL in macaques on a contrast discrimination task reduces Fano factors (***C***), noise correlations (***D***), and response variance (***F***), but has no significant effect on signal separation (***E***). We also found evidence for signal rotation (***G***) and covariance warping (***H***&***I***). The stepwise information analyses also show the similar pattern of the four mechanisms (***J***). The unit "spk/s" indicates the number of spikes per second (i.e., firing rate). We calculate aLFI and information gain using stimulus contrast as decimal values (i.e., 0.29), so they have arbitrary units. The definitions of the individual data points and error bars are the same as in Fig. 7.

DISCUSSION

In this study, we aim to answer two questions: (1) to what extent the computational theory of enhanced signal-to-noise ratio of VPL is implemented in population responses, and (2) how learning-induced changes in the activity of single neurons contribute to stimulus codes at the population level. To answer both questions, we propose a computational framework that specifies all four mechanisms by which learning may alter neural tuning and covariance structure to improve sensory discrimination. To test this framework, we first trained a DCNN model on a classical orientation discrimination task, and found that this DCNN model is able to reproduce several important empirical findings in VPL, including the downshift of TvN functions in psychophysics; enhanced multivariate decoding accuracy in human imaging; and altered tuning curves, reduced Fano factor, and reduced noise correlation in monkey neurophysiology. We confirmed the existence of three proposed mechanisms associated with learning in the DCNN model, and found that improved population codes are primarily driven by altered trial-by-trial variability of neurons, strongly supporting noise reduction as the primary mechanism of VPL. Training a DCNN model on motion direction discrimination substantiated these results. Finally, we tested these findings on multivariate BOLD activity in the human brain measured by fMRI and population spiking activity in macaque area V4 obtained via chronically implemented multi-electrode arrays. Both human and macaque data support the existence of three mechanisms underlying VPL and consistently show that noise reduction is the primary driver of improved population codes. These converging results represent a step forward in uncovering the shared computational mechanisms of VPL in both biological and artificial visual systems.

Population codes in DNNs

Our results are particularly relevant for understanding the mechanisms of perceptual learning in both deep and biological neural networks. By definition, perceptual learning refers to the process by which a perceptual system learns a task through training. Coincidentally, a similar form of learning is central to computer vision, especially when DNNs are trained to classify large numbers of natural images into a range of categories, as seen in benchmark challenges such as ImageNet. To understand the intricate mechanisms underlying the ability of these networks to discriminate between different image features, researchers have traditionally used dimensionality reduction techniques (e.g., t-SNE) to visualize and inspect the topology of feature vectors across different

image categories. To fill this gap, our proposed methodology provides a more quantifiable and interpretable framework for understanding feature representations in natural image classification. In the context of our present study, the classification task aims to only discriminate between only two stimulus values (e.g., different orientations, motion directions, or contrasts). Nevertheless, this binary approach can be easily extended to multi-class classification using strategies such as One-vs-One or One-vs-Rest. It is also worth noting that, in our study, the variability in the multidimensional distribution of image features originates from sensory noise. By contrast, in real-world image classification tasks, this variability may arise from different data augmentation techniques or from inherent variances within a single category.

Our method is also related to Local Interpretable Model-Agnostic Explanations (LIME, ref. 40). Both methods use linear approximations to understand how features can be mapped to categorization outputs. However, LIME focuses more on the pixel domain, trying to understand how and why certain segments or regions of an image contribute to a model's classification decision. Conversely, our method delves deeper into the abstract feature domain, trying to understand how and why certain feature topologies contribute to a model's classification decision. Together, they provide a comprehensive toolkit for understanding the intricate interplay of image and feature domains in model behavior.

Disentangle the effects of various factors that affect population codes

According to our results, VPL can enhance signals, reduce the response variance of individual neurons, rotate high-dimensional response distributions (i.e., signal rotation), and warp their covariance. An important contribution here is that we provide a quantitative and interpretable framework to disentangle the contributions of the four mechanisms. The signal rotation mechanism appears to dramatically enhance information due to the misalignment of covariance and signal directions (Figs. 3&4). Interestingly, this seemingly beneficial effect of signal rotation is largely counteracted by the detrimental effects of covariance warping from pre- to post-test (Figs. 3&4). After the effects of all four mechanisms are combined, the final information gain compared to baseline is still pronounced.

In our stepwise approach, we set the order of the mechanisms as signal enhancement, noise reduction, signal rotation, and covariance warping. We decompose VPL in this way because the four components can be easily visualized and are intuitively interpretable. Here, the joint effects

of signal rotation and covariance are modest (i.e., magenta bars compared to blue bars in Fig. 4&Fig. 6K&Fig. 8J, negative effects in Fig. 5L). We systematically evaluated the relationships between the four mechanisms by reversing their order in Supplementary Fig. S6. We also wanted to distinguish between noise reduction and covariance warping. In noise reduction, the total variance of the high-dimensional distributions was scaled down (i.e., $\underline{\lambda}$ in Eq. 4 is reduced), but the overall shape of the distributions remained unchanged. Thus, the two stimulus distributions simply shrunk to a "smaller size" (Fig. 4). Covariance warping redistributed the variance of high-dimensional distributions in different directions (i.e., $\hat{\lambda}_i$ and ξ_i in Eq. 4 are changed). Thus, the shape of the two stimulus distributions was warped.

Characterizing signal and noise changes during perceptual learning

Our results are consistent with the long-standing noise reduction hypothesis of VPL, which has been supported by several psychophysical studies and theoretical models [3,4,19]. The noise reduction hypothesis postulates that perceptual training improves perceptual efficiency by selectively enhancing task-relevant information and attenuating task-irrelevant noise. In our DCNN models, perceptual training only optimizes model performance by adjusting connection weights via backpropagation, without incorporating handcraft plasticity rules or other additional neural properties. It turns out that several empirical findings and the noise reduction process emerge naturally as neural consequences of such performance-optimized training.

Bejjanki et al. (2011)[41] built a biological neural network that mimics information processing from the retina to V1. Similar to our approach, the model simulates 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 (pre-trained artificial DCNNs), our study shows the convergent result that the signal separation induced by training is modest. In particular, we replicated the finding of sharpened orientation-selective tuning curves reported in the neurophysiological literature [6,7]. This phenomenon has long been hypothesized to improve population codes because uniformly reducing the tuning width of all neurons in a population with homogeneous tuning curves can significantly increase the Euclidean distance between two high-dimensional response distributions. However, this effect may be modest if the tuning curves are remarkably heterogeneous and the tuning changes vary substantially across neurons in a

population (i.e., tuning is sharpened in some neurons and broadened in others, see Fig. 2F&Fig. 5G&Fig. 6E, and ref. 33). Our modeling 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.

Our emphasis on noise reduction conflicts with some psychophysical and neurophysiological studies that support a signal enhancement mechanism 10,42. For example, some psychophysical studies have shown that signal enhancement, rather than noise reduction, is the primary mechanism of VPL 42,43. This conclusion is usually based on behavioral modeling of a uniform downshift of TvN curves. We emphasize that behavioral modeling on pure psychophysical data usually lacks sufficient details of neural implementations. Furthermore, we replicate the uniform downshift of TvN curves in our DCNN models and further show that it is due to reduced variability of neural activity.

At the neural level, Yan, et al. (2014)10 also used chronically implemented multi-electrode arrays in V1 to track neural changes when macaques were trained on a contour detection task. They analyzed the population responses and found that the signal, but not the noise, of the optimal Fisher linear discriminant changed significantly during training. The main difference between their study and our is that Yan, et al. (2014)10 computed the signal and noise of the decision unit after projecting the high-dimensional population responses onto a one-dimensional optimal decision plane, whereas we computed the signal (i.e., the Euclidean distance between two response distributions) and noise (i.e., the response variance) based on the high-dimensional population responses per se. To resolve the discrepancy, we first performed an analysis similar to that of Yan, et al. (2014)10 and replicated their results in layer 1 of the orientation DCNN model (Supplementary Fig. S4). In addition, we analytically derived that the signal and noise defined in Yan, et al. (2014)10 encompass all four mechanisms as defined in our study (Supplementary Note 1), which makes the interpretation of these individual mechanisms difficult. Mechanisms such as signal rotation and covariance warping can only be understood in multivariate responses. Here, we quantitatively disentangled the contributions of different mechanisms to population codes. The results here are more interpretable.

The effects of internal noise and signal enhancement

The highly consistent findings in both artificial and biological visual systems demonstrate the potential utility of neural network models of brain plasticity, and we highlight two aspects of neural responses that can be further explored in future work.

First, our DCNN models can only qualitatively, not quantitatively, replicate several empirical findings. An important reason is that the models do not account for the effects of internal noise in neural processing. In our DCNN model, all the stochasticity in neural responses is due to noise variation in the input images. Thus, the overall magnitude of the Fano factor derived from artificial neurons is much smaller than that reported in empirical studies. In contrast, in both human fMRI and macaque recording experiments, we did not deliberately manipulate stimulus noise, so the variability of activity was mostly driven by internal brain noise. In neurophysiological experiments, empirically measured neural variability reflects both internal and external noise. It remains unclear what the nature and magnitude of purely internal noise is. Furthermore, it is difficult to disentangle the internal and external noise in empirically measured response variability. To simplify our modeling, we decided not to make arbitrary assumptions about the nature and magnitude of internal noise, and not to even include internal noise in our DCNN modeling. Instead, we focus only on the pre-post changes in total response variability rather than disentangling the sources of response variability.

Second, although we found noise reduction as a unified mechanism across several experimental conditions, we do not claim that signal enhancement or other mechanisms plays no role in VPL. Here we examined orientation, motion, and contrast discrimination learning tasks because they have been associated with a large body of psychophysical, human imaging, and macaque neurophysiological evidence. It remains unclear whether noise reduction also acts as the primary mechanism in other brain regions, other stimulus conditions, or other perceptual learning tasks. For example, contrast detection training has been shown to significantly enhance activity in the human lateral geniculate nucleus 44. Nevertheless, our modeling and analysis methods, especially the information-theoretic analyses of population responses, provide a useful computational framework for future studies to dissect the possible contributions of signal and noise in many stimulus and task conditions.

In summary, our work combines DCNN modeling, human fMRI, and macaque multi-electrode recordings to demonstrate that VPL significantly reduces sensory noise in visual processing. These converging results across species, measurement modalities, brain regions, and

behavioral tasks suggest deep connections between the mechanisms of neural plasticity in biological and artificial visual systems.

METHODS AND MATERIALS

DCNN modeling of orientation VPL

Stimuli. The network was trained to discriminate whether a target stimulus was tilted 1º 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/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 (8 levels: 0.005, 1, 5, 10, 15, 30, 50, 75). Similar to existing psychophysical studies 45, 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 17. 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 145º) were used and we trained four DCNN models for each of the four reference orientations.

Neural network model and training. A DCNN 18 was used to simulate the orientation VPL. We retained the first five convolutional layers of the pre-trained 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 (2AFC) 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:

$$p = \frac{e^{h_t - h_r}}{1 + e^{h_t - h_r}} \tag{1}$$

The entire training procedure consisted of two distinct phases: the pre-training phase and the VPL phase. In the pre-training 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 (10 levels) and noise levels (8 levels). The network was trained for 5000 epochs in the pre-training phase and 500 epochs in the VPL phase using the stochastic gradient descent learning algorithm. The learning rate and the momentum were set to $1e^{-5}$ 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. 17, while those in the convolutional layers

were taken directly from a pre-trained AlexNet available at http://dl.caffe.berkeleyvision.org/bvlc_AlexNet.caffemodel. We trained one model for each of the four reference orientations, and the entire procedure was repeated 10 times for each reference orientation to control for randomness. All model and training procedures were implemented using PyTorch 1.13.1.

Behavioral and neural changes. The behavioral performance of the network was evaluated by measuring its classification confidence (Eq. 1) at all combinations of contrast and noise before and after the VPL stage. Specifically, in each trial, the firing rate of each artificial neuron was measured as the output of local response normalization or ReLU layers, averaged over all locations. All measurements were obtained by simulating 1000 trials for better estimation. To ensure that neurons were truly driven by the stimuli, only neurons with a mean firing rate greater than 0.001 before and after training were included in the analyses 17.

The behavioral TvN curves of the model were derived for comparison with human psychophysical results. Specifically, for each noise level, the contrast threshold was obtained by interpolating the pre- and post-test accuracies of 55% and 70% respectively. 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 1000 simulated trials, while the other half served as the test dataset.

To characterize the response properties of individual neurons, we measured orientation-selective tuning curves by sweeping the orientation of high-contrast stimuli from 0 to 180º. The tuning curves were derived by averaging 100 simulated trials for each orientation. The resulting tuning curves were then smoothed with a 10º Gaussian kernel. To control the heterogenous response range across neurons, we then normalized the tuning curves of each neuron by its maximum response and averaged the tuning curves across neurons 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 neuron, we simulated 1000 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 neuron firing rates over the 1000 simulated trials for each reference orientation. We took the median of the Fano factor across neurons 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 behavioral improvements, we applied linear Fisher information analysis to population responses. We considered the firing rates of the same groups of neurons 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 neurons 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 VPL is referred to as the signal rotation angle.

To measure how much information was contained in a layer per neuron, we calculated the average linear Fisher information (aLFI) as follows:

$$aLFI = \frac{1}{n} * \frac{df^T * \overline{\Sigma}^{-1} * df}{\Delta\theta^2} \tag{2}$$

$$\begin{aligned} \overline{\Sigma} &= \frac{\Sigma_1 + \Sigma_2}{2} \\ &= \frac{V_1^T * C_1 * V_1 + V_2^T * C_2 * V_2}{2} \end{aligned} \tag{3}$$

where n is the number of neurons in a layer, $\Delta\theta$ is the separation between the target stimulus and the reference stimulus (i.e., 1º). df is the signal vector. $\overline{\Sigma}$ is the mean of the covariance matrices (i.e., Σ_1 and Σ_2) of neurons responding to the two stimuli. V is a diagonal matrix with the variance of the neurons as the diagonal terms. 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 eigen decomposition on the covariance matrix $\overline{\Sigma}$, where we obtained λ_i, the eigenvalue of $\overline{\Sigma}$ and ξ_i, its corresponding normalized eigenvector. The aLFI can be rewritten as follows:

$$\begin{aligned} aLFI &= \frac{1}{n\Delta\theta^2} * \sum_i \frac{(df * \xi_i)^2}{\lambda_i} \\ &= \frac{1}{n\Delta\theta^2} * \frac{|df|^2}{\underline{\lambda}} * \sum_i \frac{\left(\frac{df}{|df|} * \xi_i\right)^2}{\hat{\lambda}_i} \end{aligned} \tag{4}$$

where $\underline{\lambda}$ is the mean variance, and $\lambda_i = \underline{\lambda} * \hat{\lambda}_i$. $\frac{df}{|df|}$ is the unit vector with length of 1 and direction as the same as the signal vector df. According to Eq. 4, we disentangled the potential mechanisms of improved LFI into four subparts: signal enhancement, reflected by the modulus length $|df|$; noise reduction, reflected by the mean variance of $\bar{\lambda}$; signal rotation, reflected by the direction of the signal vector df, and covariance warping, reflected by the relative angle of both ξ_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:

$$aLFI_{pre} = \frac{1}{n\Delta\theta^2} * \frac{|df_{pre}|^2}{\underline{\lambda}_{pre}} * \sum_i \frac{\left(\frac{df_{pre}}{|df_{pre}|} * \xi_i^{pre}\right)^2}{\hat{\lambda}_i^{pre}} \tag{5}$$

Only considering the effect of signal enhancement, we can calculate its effect as $aLFI_{se}$:

$$aLFI_{se} = \frac{1}{n\Delta\theta^2} * \frac{|df_{post}|^2}{\underline{\lambda}_{pre}} * \sum_i \frac{\left(\frac{df_{pre}}{|df_{pre}|} * \xi_i^{pre}\right)^2}{\hat{\lambda}_i^{pre}} \tag{6}$$

Note that the only difference here is that the $|df_{pre}|^2$ in Eq. 5 is replaced by the $|df_{post}|^2$ in Eq. 6. The difference between $aLFI_{se}$ and $aLFI_{pre}$ is considered as the information gain introduced by the signal enhancement mechanism (i.e., the brown bars in in Fig. 4B). Following this idea, we can calculate the stepwise $aLFI$ by one-by-one considering the effects of noise reduction $aLFI_{nr}$, signal rotation $aLFI_{sr}$, and covariance warping ($aLFI_{cw}$ or $aLFI_{post}$) as:

$$aLFI_{nr} = \frac{1}{n\Delta\theta^2} * \frac{|df_{post}|^2}{\underline{\lambda}_{post}} * \sum_i \frac{\left(\frac{df_{pre}}{|df_{pre}|} * \xi_i^{pre}\right)^2}{\hat{\lambda}_i^{pre}} \tag{7}$$

$$aLFI_{sr} = \frac{1}{n\Delta\theta^2} * \frac{|df_{post}|^2}{\underline{\lambda}_{post}} * \sum_i \frac{\left(\frac{df_{post}}{|df_{post}|} * \xi_i^{pre}\right)^2}{\hat{\lambda}_i^{pre}} \tag{8}$$

$$aLFI_{cw} = aLFI_{post} = \frac{1}{n\Delta\theta^2} * \frac{|df_{post}|^2}{\underline{\lambda}_{post}} * \sum_i \frac{\left(\frac{df_{post}}{|df_{post}|} * \xi_i^{post}\right)^2}{\hat{\lambda}_i^{post}} \tag{9}$$

The information gain in Fig. 4B indicates the difference between $aLFI_{se}$, $aLFI_{nr}$, $aLFI_{sr}$, and $aLFI_{cw}$ (i.e., $aLFI_{post}$) as compared to the pre-test baseline $aLFI_{pre}$. They are shown as brown, blue, green, and magenta bars in Fig. 4B, respectively.

765 ***DCNN modeling of motion VPL***

766 *Stimuli*. The experiment used random dot motion (RDM) stimuli, which consist of a cloud
767 of independent moving dots with some degree of coherence in a given moving direction 46. The
768 network was trained to discriminate whether the moving direction of a target RDM stimulus was
769 4º clockwise or counterclockwise relative to its corresponding reference RDM stimulus. To meet
770 the network's specifications, the motion stimuli were 16-frame videos (112 × 112 pixels per frame).
771 Within each frame, ~100 dots were displayed, with each dot represented by a cross of 3 pixels in
772 both height and width. We set 8 coherence levels (8.84, 12.5, 17.7, 25, 35.3, 50, 70.7, and 100%)
773 and four reference directions (45, 135, 225, and 315º). The motion speed was 7.5 pixels/frame. All
774 non-coherently moving dots appeared randomly in the image. The display of each frame was
775 limited to a centered circle with a diameter of 112 pixels, with the surrounding areas displayed in
776 black.

777 *Neural network architecture and training*. Our DCNN model is a 3D convolutional neural
778 network inherited from the C3D network for action recognition 37. The original C3D consists of
779 10 convolutional layers and 3 fully connected layers. The main difference between C3D and
780 AlexNet is that C3D uses 3D convolutional kernels to process spatiotemporal information. We
781 kept the first 6 convolutional layers from the pre-trained C3D and replaced the 3 fully connected
782 layers with a fully connected layer that outputs a single scalar. The number of layers was chosen
783 to (1) keep roughly similar number of parameters to the orientation DCNN, and (2) to roughly
784 match the number of ROIs in the human neuroimaging experiment. Similar to the orientation
785 DCNN model, the motion DCNN model was also configured in a Siamese fashion to perform the
786 2AFC task based on the sigmoid function.

787 Similar to the orientation DCNN model, the entire training procedure consisted of two
788 phases: the pre-training phase and the VPL phase. During the pre-training phase, the network was
789 trained on full-coherence noiseless RDM pairs, while during the VPL phase, the network was
790 trained on stimulus pairs across all coherence levels (8 levels). The network was trained for 1000
791 epochs in the pre-training phase and 2000 epochs in the VPL phase using stochastic gradient
792 descent (SGD) with a learning rate of $1e^{-7}$, the momentum of 0.9, and a weight decay of 0.0005.
793 The parameters were updated to minimize the cross-entropy loss between the network outputs and
794 the true stimulus labels. The initial parameters in the fully connected layer were normally
795 randomized while those in the convolutional layers were taken directly from a pre-trained C3D

796 available at
797 https://download.openmmlab.com/mmaction/recognition/c3d/c3d_sports1m_16x1x1_45e_ucf10
798 1_rgb/c3d_sports1m_16x1x1_45e_ucf101_rgb_20201021-26655025.pth. The entire procedure
799 was repeated 10 times for each reference direction to control for randomness.

800 *Behavioral and neural analyses.* The behavioral performance of the network was also
801 evaluated by its classification confidence (Eq. 1) at all coherence levels before and after the VPL
802 phase. In addition, the firing rates of artificial neurons were measured on each trial as the output
803 of the ReLU layers, averaged over all locations and time points. All measurements were taken over
804 1000 simulated trials. To ensure that neurons were truly driven by the stimuli, only neurons with
805 a mean firing rate greater than 0.001 before and after training were included in subsequent analyses.

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

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

819 Again, we computed linear Fisher information using a stepwise approach. For all models,
820 we sequentially added signal enhancement, noise reduction, signal rotation, and covariance
821 warping to the calculation of linear Fisher information and see how the information within neurons
822 varies with all four mechanisms. Fig. 5L shows the results of the stepwise analysis in layer 6.
823 Supplementary Fig. S3 shows the results in all six layers of the motion DCNN model.

824

825 ***Human fMRI experiment***

826 The human fMRI experiment data have been published in ref. 38 for different research questions.
827 The core analyses in this study beyond preprocessing and ROI definitions are novel. We provide
828 relevant methods as follows and more detailed methods in Supplementary Note 2 to avoid cross-
829 referencing.

830 *Subjects and Experimental procedures*. A total of 22 human subjects (10 males and 12
831 females, ages 17-25) participated in the experiment. All participants had normal or correct-to-
832 normal vision. None of the participants were aware of the study's objectives. All participants
833 provided written informed consent, and the study obtained approval from the local ethics
834 committee at Peking University.

835 All subjects were trained on a direction discrimination task (Fig. 6A, see Supplementary
836 Note 2 for apparatus and stimulus details). The whole experiment consisted of three phases: pre-
837 test (2 days), training (10 days) and post-test (2 days). On day 1 at pre-test and day 2 at post-test,
838 subjects were tested on direction discrimination around 45º and 135º (angular difference 4º, 120
839 trials for each direction) to assess their behavioral performance before and after training. Subjects
840 were trained on the fine direction discrimination task for 10 days. Half of the subjects were trained
841 at 45º and the other half at 135º (see training details in Supplementary Note 2). Training-induced
842 behavioral improvements have been reported in our previous work 38.

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

848 *MRI data acquisition.* All MRI data were acquired using a 12-channel phase array coil on
849 a Siemens Trio 3T scanner at Peking University. The T1-weighted anatomical data with a
850 resolution of $1 \times 1 \times 1$ mm³ were collected for each subject. Echo-planar imaging (EPI) functional
851 data were collected for the motion direction discrimination task, retinotopic mapping, and motion
852 localizer experiments. EPI data were acquired using gradient echo-pulse sequences from 33 axial
853 slices, covering the whole brain. The standard EPI sequence used for data acquisition was as
854 follows: a repetition time of 2000 ms, an echo time of 30 ms, a flip angle of 90º, and a resolution
855 of $3 \times 3 \times 3$ mm³. 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 52,55 and a motion localizer run 47 to define regions-of-interest.

MRI data analyses. In Brain Voyager QX, 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 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 48,49. 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 (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 MNI coordinate space using an EPI template. We used the GLMdenoise package developed in ref. 50 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 signal-to-noise ratio, 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, e.g., 45º vs. untrained direction, e.g., 135º) 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 45º vs. 135º conditions to quantify stimulus discriminability. We found that motion VPL significantly improved stimulus discriminability in V3A and hMT+. Therefore, we only included V3A and hMT+ voxels in the subsequently analyses.

886 Similar to the analyses in the DCNN models, we defined the signal vector, the signal
887 separation, the variance, the intervoxel correlations, the signal rotation angle, the PC strength, and
888 the PC rotation angle in the multivoxel high-dimensional space using the same method defined
889 above (Fig. 6). In addition, we applied the same stepwise analysis approach of calculating aLFI to
890 the fMRI data (Fig. 6K).
891
892 ***Monkey multi-electrode recording experiment***
893 Part of the monkey psychophysical and neurophysiological data have been published in refs. 11,51.
894 These previous studies showed qualitatively similar results of the learning-induced reduction in
895 Fisher information, Fano factor, and noise correlations via different analysis methods. Other results
896 and analyses on the characteristics of population responses in this study (i.e., Figs. 7&8), especially
897 the validation of signal rotation and covariance warping mechanisms, as well as the stepwise
898 information analyses, are novel. We provide relevant methods as follows and more detailed
899 methods in Supplementary Note 3 to avoid cross-referencing.
900 *Ethics statement and Data collection*. The Newcastle University Animal Welfare Ethical
901 Review Board approved all procedures in this study. All experimental procedures were carried out
902 in accordance with the European Communities Council Directive RL 2010/63/EC, the US National
903 Institutes of Health Guidelines for the Care and Use of Animals for Experimental Procedures and
904 the UK Animals Scientific Procedures Act. This study included two male macaque monkeys (5
905 and 14 years of age).
906 *Experimental preparation*. The surgical procedure is described in ref. 52 and
907 Supplementary Note 3. The headpost and electrode implementations are also described in
908 Supplementary Note 3. Briefly, in monkey 1, two 4 × 5 grids of microelectrodes were implanted
909 in area V4; in monkey 2, one 5 × 5 grid was implanted in V4. These chronically implanted
910 electrodes allowed us to record population activity in area V4 over the course of perceptual training.
911 Importantly, we were able to record stably from a few small multi-unit clusters. The stability of
912 the recording is shown in ref. 11. Stable recording of multi-channel neuronal activity allows
913 analyses of changes in population responses induced by training.
914 *Behavioral task and monkey training*. The monkeys were trained in a contrast
915 discrimination task in which subjects were asked to decide whether the contrast of a test stimulus
916 was higher or lower as compared to that of a reference stimulus by making a saccade to one of two

917 distinct locations (Fig. 7B). On each trial, the subject first kept fixation on the center-of-screen for
918 512ms. After 539 ms of fixation, a vertically oriented reference Gabor stimulus with 30% contrast
919 was presented, centered at the V4 Receptive Field (RF) coordinates. The outer diameter of the
920 Gabor stimulus was truncated at 16° for monkey 1 and 14° for monkey 2. Following the Gabor
921 stimulus, monkey 2 experienced an interstimulus interval of 512 ms. In contrast, monkey 1
922 experienced a randomly chosen interstimulus interval, ranging from 512 to 1024 ms. During the
923 interstimulus interval, only the fixation dot was presented. A test stimulus was then presented for
924 512 ms. This test stimulus was identical in size and orientation to the reference stimulus but
925 differed in contrast, with the contrast level chosen pseudorandomly. The test stimulus was
926 followed by another blank period of 512 ms during which only the fixation dot was visible. After
927 the fixation dot, two target squares, one black and one white with a size of 0.5° in size, appeared
928 to the left and right of the location where the reference and test stimuli were previously presented.
929 The monkeys were cued to make a decision once the fixation dot disappeared. The monkeys were
930 required to make a saccade to the white square within a 2° × 2° window if the test stimulus had a
931 higher contrast than the reference stimulus. Conversely, they were expected to make a saccade to
932 the black square if the test stimulus had a lower contrast than the reference stimuli. A correct
933 saccade was rewarded with a fluid reward, while an incorrect saccade led to no reward and a 0.2 s
934 timeout period.

935 The two monkeys were first trained on an easy version (target contrast 5% or 90%) of the
936 contrast discrimination task. After they were fully familiar with the easy task, the target contrast
937 increased from 2 to 8, 12, and 14 levels. The data correspond to the 14 levels of target contrast (10,
938 15, 20, 25, 27, 28, 29, 31, 32, 33, 35, 40, 50 or 60%; Supplementary Note 3). We only focus on
939 target contrast levels (27, 28, 29, 31, 32, and 33%) near the reference contrast (i.e., 30%) according
940 to the definition of linear Fisher information.

941 *Dataset and preprocessing*. We used chronically implanted Utah arrays to record spiking
942 activity. We refer to small multi-unit neuronal clusters recorded from a given electrode as
943 'channels'. Twenty-nine and Twenty channels were recorded in monkey 1 and monkey 2,
944 respectively. These channels exhibited good responses (signal-to-noise ratio [SNR] >1) on over
945 80% of the recording sessions (see SNR computation in Supplementary Note 3). Baseline activity
946 matching was performed between sessions for multi-unit activity (MUA) data in order to obtain
947 comparable activity levels across sessions.

Behavioral and neural analyses. We noticed that the relationship between neural activity and discriminability can change drastically during the stimulus presentation period, and through perceptual learning, the improvement in discriminability can also vary over the course of the training period. We chose the first 4 and the last 4 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 10-fold 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 codes (similar to the decoding analyses for first identifying V3A and hMT+ 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 noise reduction. Also, varying the time window did not qualitatively change our results. We used a simple multivariate Poisson log-normal (MPLN) model (Supplementary Note 3, see also refs. 53-56) 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&8 for visual comparison with the DCNN and fMRI results above.

CONFLICT OF INTERESTS

XC is a cofounder and shareholder of a neurotechnology start-up, Phosphoenix (Netherlands). Other authors declare no competing financial interests.

ACKNOWLEDGEMENT

We thank Shuguang Kuai, Duje Tadin, Oh-Sang Kwon, and Zhong-Lin Lu for valuable comments on the manuscripts. This work was supported by the National Natural Science Foundation of China (32100901), Natural Science Foundation of Shanghai (21ZR1434700), the Research Project of

Shanghai Science and Technology Commission (20dz2260300) and the Fundamental Research Funds for the Central Universities (to R.-Y.Z.). The macaque work was supported by the Medical Research Council, UK, G0700976.

AUTHOR CONTRIBUTIONS

Y-A. C. and R-Y.Z. conceived and designed the study. Y-A. C. implemented the neural network modeling. J.K., 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. and R-Y.Z. wrote the first draft of the manuscript. All authors revised the manuscript and provided valuable feedback to the final manuscript.

RFERENCES

1. Lu, Z.L., Yu, C., Watanabe, T., Sagi, D., and Levi, D. (2010). Perceptual learning: functions, mechanisms, and applications. Vision Res *50*, 365-367. 10.1016/j.visres.2010.01.010.
2. Watanabe, T., and Sasaki, Y. (2015). Perceptual learning: toward a comprehensive theory. Annu Rev Psychol *66*, 197-221. 10.1146/annurev-psych-010814-015214.
3. Lu, Z.L., and Dosher, B.A. (2008). Characterizing observers using external noise and observer models: assessing internal representations with external noise. Psychol Rev *115*, 44-82. 10.1037/0033-295X.115.1.44.
4. Dosher, B.A., and Lu, Z.L. (1998). Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. Proc Natl Acad Sci U S A *95*, 13988-13993. 10.1073/pnas.95.23.13988.
5. Dosher, B., and Lu, Z.L. (2017). Visual Perceptual Learning and Models. Annu Rev Vis Sci *3*, 343-363. 10.1146/annurev-vision-102016-061249.
6. Schoups, A., Vogels, R., Qian, N., and Orban, G. (2001). Practising orientation identification improves orientation coding in V1 neurons. Nature *412*, 549-553. 10.1038/35087601.
7. Yang, T., and Maunsell, J.H. (2004). The effect of perceptual learning on neuronal responses in monkey visual area V4. J Neurosci *24*, 1617-1626. 10.1523/JNEUROSCI.4442-03.2004.
8. Ni, A.M., Ruff, D.A., Alberts, J.J., Symmonds, J., and Cohen, M.R. (2018). Learning and attention reveal a general relationship between population activity and behavior. Science *359*, 463-465. 10.1126/science.aao0284.
9. Gu, Y., Liu, S., Fetsch, C.R., Yang, Y., Fok, S., Sunkara, A., DeAngelis, G.C., and Angelaki, D.E. (2011). Perceptual learning reduces interneuronal correlations in macaque visual cortex. Neuron *71*, 750-761. 10.1016/j.neuron.2011.06.015.
10. Yan, Y., Rasch, M.J., Chen, M., Xiang, X., Huang, M., Wu, S., and Li, W. (2014). Perceptual training continuously refines neuronal population codes in primary visual cortex. Nat Neurosci *17*, 1380-1387. 10.1038/nn.3805.
11. Sanayei, M., Chen, X., Chicharro, D., Distler, C., Panzeri, S., and Thiele, A. (2018). Perceptual learning of fine contrast discrimination changes neuronal tuning and population coding in macaque V4. Nat Commun *9*, 4238. 10.1038/s41467-018-06698-w.
12. Chen, N., Bi, T., Zhou, T., Li, S., Liu, Z., and Fang, F. (2015). Sharpened cortical tuning and enhanced cortico-cortical communication contribute to the long-term neural mechanisms of visual motion perceptual learning. Neuroimage *115*, 17-29. 10.1016/j.neuroimage.2015.04.041.
13. Shibata, K., Chang, L.H., Kim, D., Nanez, J.E., Sr., Kamitani, Y., Watanabe, T., and Sasaki, Y. (2012). Decoding reveals plasticity in V3A as a result of motion perceptual learning. PLoS One *7*, e44003. 10.1371/journal.pone.0044003.
14. Jia, K., Zamboni, E., Kemper, V., Rua, C., Goncalves, N.R., Ng, A.K.T., Rodgers, C.T., Williams, G., Goebel, R., and Kourtzi, Z. (2020). Recurrent Processing Drives Perceptual Plasticity. Curr Biol *30*, 4177-4187 e4174. 10.1016/j.cub.2020.08.016.
15. Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., and Pouget, A. (2014). Information-limiting correlations. Nat Neurosci *17*, 1410-1417. 10.1038/nn.3807.

16. Gilpin, L.H., Bau, D., Yuan, B.Z., Bajwa, A., Specter, M., and Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. (IEEE), pp. 80-89.
17. Wenliang, L.K., and Seitz, A.R. (2018). Deep Neural Networks for Modeling Visual Perceptual Learning. J Neurosci *38*, 6028-6044. 10.1523/JNEUROSCI.1620-17.2018.
18. Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems *25*, 1097-1105.
19. Dosher, B.A., and Lu, Z.L. (1999). Mechanisms of perceptual learning. Vision Res *39*, 3197-3221. 10.1016/s0042-6989(99)00059-0.
20. Chen, N., Cai, P., Zhou, T., Thompson, B., and Fang, F. (2016). Perceptual learning modifies the functional specializations of visual cortical areas. Proc Natl Acad Sci U S A *113*, 5724-5729. 10.1073/pnas.1524160113.
21. Raiguel, S., Vogels, R., Mysore, S.G., and Orban, G.A. (2006). Learning to see the difference specifically alters the most informative V4 neurons. J Neurosci *26*, 6589-6602. 10.1523/JNEUROSCI.0457-06.2006.
22. Ghose, G.M., Yang, T., and Maunsell, J.H. (2002). Physiological correlates of perceptual learning in monkey V1 and V2. J Neurophysiol *87*, 1867-1888. 10.1152/jn.00690.2001.
23. Jehee, J.F., Ling, S., Swisher, J.D., van Bergen, R.S., and Tong, F. (2012). Perceptual learning selectively refines orientation representations in early visual cortex. J Neurosci *32*, 16747-16753a. 10.1523/JNEUROSCI.6112-11.2012.
24. Kohn, A., Coen-Cagli, R., Kanitscheider, I., and Pouget, A. (2016). Correlations and Neuronal Population Information. Annu Rev Neurosci *39*, 237-256. 10.1146/annurev-neuro-070815-013851.
25. Jeanne, J.M., Sharpee, T.O., and Gentner, T.Q. (2013). Associative learning enhances population coding by inverting interneuronal correlation patterns. Neuron *78*, 352-363. 10.1016/j.neuron.2013.02.023.
26. Ruff, D.A., and Cohen, M.R. (2014). Attention can either increase or decrease spike count correlations in visual cortex. Nat Neurosci *17*, 1591-1597. 10.1038/nn.3835.
27. Hennequin, G., Ahmadian, Y., Rubin, D.B., Lengyel, M., and Miller, K.D. (2018). The Dynamical Regime of Sensory Cortex: Stable Dynamics around a Single Stimulus-Tuned Attractor Account for Patterns of Noise Variability. Neuron *98*, 846-860 e845. 10.1016/j.neuron.2018.04.017.
28. Ecker, A.S., Berens, P., Keliris, G.A., Bethge, M., Logothetis, N.K., and Tolias, A.S. (2010). Decorrelated neuronal firing in cortical microcircuits. Science *327*, 584-587. 10.1126/science.1179867.
29. Gutnisky, D.A., and Dragoi, V. (2008). Adaptive coding of visual information in neural populations. Nature *452*, 220-224. 10.1038/nature06563.
30. Huang, X., and Lisberger, S.G. (2009). 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. 10.1152/jn.00010.2009.
31. van Bergen, R.S., and Jehee, J.F.M. (2018). Modeling correlated noise is necessary to decode uncertainty. Neuroimage *180*, 78-87. 10.1016/j.neuroimage.2017.08.015.
32. Cheng, Z.J., Yang, L., Zhang, W.H., and Zhang, R.Y. (2023). Representational Geometries Reveal Differential Effects of Response Correlations on Population Codes in

Neurophysiology and Functional Magnetic Resonance Imaging. J Neurosci *43*, 4498-4512. 10.1523/JNEUROSCI.2228-22.2023.

33. Zhang, R.Y., Wei, X.X., and Kay, K. (2020). Understanding multivariate brain activity: Evaluating the effect of voxelwise noise correlations on population codes in functional magnetic resonance imaging. PLOS Computational Biology *16*, e1008153. 10.1371/journal.pcbi.1008153.
34. Liu, Z. (1999). Perceptual learning in motion discrimination that generalizes across motion directions. Proc Natl Acad Sci U S A *96*, 14085-14087. 10.1073/pnas.96.24.14085.
35. Ball, K., and Sekuler, R. (1987). Direction-specific improvement in motion discrimination. Vision Res *27*, 953-965. 10.1016/0042-6989(87)90011-3.
36. Law, C.T., and Gold, J.I. (2008). Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. Nat Neurosci *11*, 505-513. 10.1038/nn2070.
37. Tran, D., Bourdev, L., Fergus, R., Torresani, L., and Paluri, M. (2015). Learning spatiotemporal features with 3d convolutional networks. pp. 4489-4497.
38. Jia, K., Xue, X., Lee, J.H., Fang, F., Zhang, J., and Li, S. (2018). Visual perceptual learning modulates decision network in the human brain: The evidence from psychophysics, modeling, and functional magnetic resonance imaging. J Vis *18*, 9. 10.1167/18.12.9.
39. Thomson, H.L., O'Brien, M.F., Almeida, A.A., Tesar, P.J., Davison, M.B., and Burstow, D.J. (1998). Haemodynamics and left ventricular mass regression: a comparison of the stentless, stented and mechanical aortic valve replacement. Eur J Cardiothorac Surg *13*, 572-575. 10.1016/s1010-7940(98)00058-x.
40. Ribeiro, M.T., Singh, S., and Guestrin, C. (2016). " Why should i trust you?" Explaining the predictions of any classifier. pp. 1135-1144.
41. Bejjanki, V.R., Beck, J.M., Lu, Z.L., and Pouget, A. (2011). Perceptual learning as improved probabilistic inference in early sensory areas. Nat Neurosci *14*, 642-648. 10.1038/nn.2796.
42. Gold, J., Bennett, P.J., and Sekuler, A.B. (1999). Signal but not noise changes with perceptual learning. Nature *402*, 176-178. 10.1038/46027.
43. Poggio, T., Fahle, M., and Edelman, S. (1992). Fast perceptual learning in visual hyperacuity. Science *256*, 1018-1021. 10.1126/science.1589770.
44. Yu, Q., Zhang, P., Qiu, J., and Fang, F. (2016). Perceptual Learning of Contrast Detection in the Human Lateral Geniculate Nucleus. Curr Biol *26*, 3176-3182. 10.1016/j.cub.2016.09.034.
45. Lu, Z.L., Chu, W., and Dosher, B.A. (2006). Perceptual learning of motion direction discrimination in fovea: separable mechanisms. Vision Res *46*, 2315-2327. 10.1016/j.visres.2006.01.012.
46. Britten, K.H., Shadlen, M.N., Newsome, W.T., and Movshon, J.A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. J Neurosci *12*, 4745-4765. 10.1523/JNEUROSCI.12-12-04745.1992.
47. Huk, A.C., Dougherty, R.F., and Heeger, D.J. (2002). Retinotopy and functional subdivision of human areas MT and MST. J Neurosci *22*, 7195-7205. 10.1523/JNEUROSCI.22-16-07195.2002.

48. Engel, S.A., Glover, G.H., and Wandell, B.A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cereb Cortex *7*, 181-192. 10.1093/cercor/7.2.181.
49. Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., and Tootell, R.B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science *268*, 889-893. 10.1126/science.7754376.
50. Kay, K.N., Rokem, A., Winawer, J., Dougherty, R.F., and Wandell, B.A. (2013). GLMdenoise: a fast, automated technique for denoising task-based fMRI data. Front Neurosci *7*, 247. 10.3389/fnins.2013.00247.
51. Chen, X., Sanayei, M., and Thiele, A. (2013). Perceptual learning of contrast discrimination in macaca mulatta. J Vis *13*, 22. 10.1167/13.13.22.
52. Thiele, A., Delicato, L.S., Roberts, M.J., and Gieselmann, M.A. (2006). 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. 10.1016/j.jneumeth.2006.05.032.
53. Goris, R.L., Movshon, J.A., and Simoncelli, E.P. (2014). Partitioning neuronal variability. Nat Neurosci *17*, 858-865. 10.1038/nn.3711.
54. Henaff, O.J., Bai, Y., Charlton, J.A., Nauhaus, I., Simoncelli, E.P., and Goris, R.L.T. (2021). Primary visual cortex straightens natural video trajectories. Nat Commun *12*, 5982. 10.1038/s41467-021-25939-z.
55. Rabinowitz, N.C., Goris, R.L., Cohen, M., and Simoncelli, E.P. (2015). Attention stabilizes the shared gain of V4 populations. Elife *4*, e08998. 10.7554/eLife.08998.
56. Zhu, R.J.B., and Wei, X.X. (2023). Unsupervised approach to decomposing neural tuning variability. Nat Commun *14*, 2298. 10.1038/s41467-023-37982-z.