Unsupervised learning as a computational principle works in visual learning of natural scenes, but not of artificial stimuli

Takeo Watanabe[1], Yuka Sasaki[1], Daiki Ogawa[2], & Kazuhisa Shibata[3]

1 Department of Cognitive and Psychological Sciences, Brown University

2 Department of Literature, Nagoya University

3 Center of Brain Science, RIKEN

Abstract

The question of whether we learn exposed visual features remains a subject of controversy. A prevalent computational model suggests that visual features frequently exposed to observers in natural environments are likely to be learned. However, this unsupervised learning model appears to be contradicted by the significant body of experimental results with human participants that indicates visual perceptual learning (VPL) of visible task-irrelevant features does not occur with frequent exposure. Here, we demonstrate a resolution to this controversy with a new finding: Exposure to a dominant global orientation as task-irrelevant leads to VPL of the orientation, particularly when the orientation is derived from natural scene images, whereas VPL did not occur with artificial images even with matched distributions of local orientations and spatial frequencies to natural scene images. Further investigation revealed that this disparity arises from the presence of higher-order statistics derived from natural scene images—global structures such as correlations between different local orientation and spatial frequency channels. Moreover, behavioral and neuroimaging results indicate that the dominant orientation from these higher-order statistics undergoes less attentional suppression than that from artificial images, which may facilitate VPL. Our results contribute to resolving the controversy by affirming the validity of unsupervised learning models for natural scenes but not for artificial stimuli. They challenge the assumption that VPL occurring in everyday life can be predicted by laws governing VPL for conventionally used artificial stimuli.

Introduction

Unsupervised learning, a fundamental principle in computational learning, posits that frequent occurrences drive learning, regardless of their relevance to the task at hand (Barlow, 1989; Doya, 1999; Glaser et al., 2019). This concept, rooted in visual processing, highlights the importance of exposure in shaping perception (Krotov & Hopfield, 2019; Matteucci et al., 2024; von der Malsburg, 1973, 1994, 1999, 2018, 2021; Willshaw et al., 1997).

However, while unsupervised learning holds sway in computational models, it is not regarded as a suitable principle for biological visual perceptual learning (VPL), known as long-term performance enhancement as a result of visual experience (Lu & Dosher, 2022; Seitz & Dinse, 2007; Shibata et al., 2014; Watanabe & Sasaki, 2015). This is because the unsupervised learning seems contradicted by a large volume of empirical evidence that VPL does not occur for visual features that are exposed but irrelevant to a given task, particularly if these features are supra-threshold and visible (Ahissar & Hochstein, 1993; Astorga et al., 2022; Huang et al., 2007; Lewis et al., 2009; Li et al., 2004; Mukai et al., 2011; Paffen et al., 2008; Schoups et al., 2001; Shiu & Pashler, 1992; Tsushima et al., 2008). This lack of VPL has been attributed to attentional suppression of exposed irrelevant features for efficient task performance (Astorga et al., 2022; Sasaki et al., 2010; Watanabe & Sasaki, 2015). The findings that VPL of task-irrelevant features occurs exclusively when the features were sub-threshold (Chang et al., 2014; Frank et al., 2020; Seitz & Watanabe, 2003; Tsushima et al., 2006, 2008; Watanabe et al., 2001) further supports the theory which suggests that attentional suppression of supra-threshold task-irrelevant features occurs to eliminate conspicuous distractors from tasks, leading to no VPL (Sasaki et al., 2010; Watanabe & Sasaki, 2015).

As in the vast majority of studies of VPL (Li, 2016; Lu & Dosher, 2022; Sagi & Tanne, 1994; Sasaki et al., 2010; Seitz & Dinse, 2007; Shibata et al., 2014; Watanabe & Sasaki, 2015), all studies investigating VPL of task-irrelevant features utilized artificial stimuli rather than natural scene stimuli. Utilizing artificial stimuli such as Gabor patches (Dosher & Lu, 1999), sinusoidal gratings (Gutnisky et al., 2009), and coherent motion stimuli (Seitz & Watanabe, 2003; Tsushima et al., 2008; Wang et al., 2013; Watanabe et al., 2001) in VPL studies offers certain advantages: The parameters of artificial stimuli can be precisely controlled, facilitating experimental precision. Consistency in using the same type of artificial stimuli across studies enables the comparison and integration of results more effectively. However, simple stimuli may lack variation in spatial or temporal global structures, which are present in natural scene stimuli. Therefore, if these structures play a critical role in VPL for natural scene stimuli, it may not necessarily be predicted by VPL studies conducted with artificial stimuli.

In the current study, we examined whether exposure to a dominant supra-threshold orientation as task-irrelevant leads to VPL of the orientation in both artificial and natural scene images that have similar orientation and spatial frequency distributions using identical procedures. We found that exposure to natural scene images resulted in VPL of the dominant orientation in these images. Conversely, exposure to artificial images, despite matching local orientation and spatial frequency distribution with natural scene images, did not lead to VPL, as has been found in previous studies using artificial supra-threshold stimuli (Astorga et al., 2022; Chang et al., 2014; Frank et al., 2020; Tsushima et al., 2008). These findings suggest that the occurrence of exposure-based task-irrelevant learning depends on whether the exposed stimuli were natural scenes or artificial images.

We explored the reasons for these differential outcomes. We found that higher-order statistics derived from natural scene images, which include correlations between orientation and spatial frequency channels, are sufficient to induce VPL of task-irrelevant supra-threshold features. Furthermore, the results of functional magnetic resonance imaging (fMRI) and behavioral measurements showed that the degree of attentional suppression on these higher-order statistics derived from natural scenes is significantly lower in the visual areas known to respond to such statistics (Freeman et al., 2013; Hatanaka et al., 2022; Okazawa et al., 2015).

These results suggest that exposure to stimuli containing higher-order statistic signals derived from natural scene images leads to VPL of the dominant orientation in these stimuli, maybe because the global orientation representation generated from higher-order statistic signals is subject to weak or no attentional suppression. In contrast, the dominant orientation in artificial stimuli conventionally used for VPL studies is not generated from higher-order statistics and may be subject to significantly stronger attentional suppression than that generated from higher-order statistics derived from natural scene images.

The conclusion drawn from these results suggests that contrary to the conclusions of previous studies, the supervised learning principle is indeed effective in VPL when an exposed task-irrelevant feature is derived from a natural scene. Most studies on VPL have relied on artificial stimuli lacking these higher-order statistics. The current study suggests that VPL of stimuli typically used in research may not always generalize to VPL of natural scene stimuli, which are encountered in everyday life. Therefore, rules established through training with artificial stimuli may require reassessment when applied to VPL for natural scenes.

Results

Task-irrelevant VPL occurs based on natural scene images, but not on artificial images

In Experiment 1 (see Stimuli and Experiment 1 in Methods for details), we examined whether VPL of the dominant orientation in a task-irrelevant and supra-threshold natural scene (NS) image occurs in distinct ways from that in a supra-threshold artificial image. Each of two different groups of 12 participants were exposed to task-irrelevant NS images (NS group) or artificial images consisting of the same spectra of orientations and spatial frequencies as NS images, termed Fourier-scrambled (FS) images (FS group).

The complete procedure with both NS and FS groups consisted of one day of the pre-test stage, 10 days of the exposure stage, and one day of the post-test stage (Fig. 1a).

In the exposure stage, participants were asked to perform a rapid serial visual presentation (RSVP) task, designed to make participants' attention focused on the central task (Chang et al., 2014; Frank et al., 2020; Nishina et al., 2007; Seitz & Watanabe, 2003; Tsushima et al., 2008; Watanabe et al., 2001). In each trial, a sequence of eight alphabetical letters and two digits were presented in a random order for each participant at the center of the display (Fig. 1b). After the presentation of the sequence, the participants were instructed only to report the two digits (see Fig. S1a for performance in the RSVP task). While the RSVP task was performed, an NS (Fig. 1c, highlighted with orange rings) or FS (Fig. 1c, highlighted with green rings) image, as task-irrelevant, was being presented in the background to the NS or FS group, respectively. We created two sets of 40 NS images, each set exhibiting the dominant orientation power peaking at 45° or 135°, referred to as 45°-dominant and 135°-dominant NS images, respectively (Fig. S2). Perceiving the dominant orientation in NS is ecologically important for identifying or categorizing the NS (Dragoi et al., 2001; Girshick et al., 2011). Importantly, the FS images consisted of the identical spectra of orientations and spatial frequencies and the mean luminance and variance of luminance to those of the NS images used for the NS group (see Stimuli in Methods for details of creating the FS images). In each of the NS and FS groups, half of participants were exposed to the 45°-dominant images and the other half to the 135°-dominant images. The dominant orientation in the NS and FS images had been proven to be supra-threshold and almost equally visible (see Control Experiment in Methods and Fig. S3).

The pre-test and post-test stages examined how exposure to the NS and FS images influenced the discriminability of the dominant orientations in the presented task-irrelevant images. We used Gabor patches in both the pre- and post-test stages for both the NS and FS groups. All participants were asked to perform a discrimination task (Fig. 1d) for the dominant orientation to which they were exposed during the exposure stage (exposed orientation) and the orientation orthogonal to the dominant orientation (unexposed orientation). In each trial, participants were asked to report whether the orientations of the two Gabor patches presented successively were

the same or different. In half the trials, one Gabor patch was at the reference orientation, and the other patch was at an orientation differing by ±Δ degrees from the reference. In the other half, both Gabor patches were at the reference orientation. The reference orientation for each block was either exposed or unexposed orientation and alternated between blocks. The Δ value for each block were 12, 8, 5, 3, or 1 degree, decreasing every two blocks consisting of 48 trials. Performance in each of the pre- and post-stages was defined as the mean accuracy averaged across the Δ values of 12, 8, 5, 3, and 1 degree. An improvement in performance was defined by the subtraction of the performance in the pre-test stage from that in the post-test stage.

Fig. 1e shows improvements in the NS and FS groups (see Fig. S1b and c for task accuracies before the averaging and subtraction). We applied three-way ANOVA on the performance with Group (NS vs FS) as the between-participant factor and Test (pre vs post) and Orientation (exposed vs unexposed) as the within-participant factors (see Table S1 for detailed results of ANOVA). Since there was a significant three-way interaction ($F_{1,22}$ = 5.8216, P = 0.0246, partial η^2 = 0.2092, 95% CI of partial η^2 = [0.0036 0.4781]), we proceeded to a two-way ANOVA with the factors of Test and Orientation to explore the source of the interaction in each group (See Statistics in Methods for details). In the NS group, the results of ANOVA showed significant effects of Test ($F_{1,11}$ = 9.2612, P = 0.0112, partial η^2 = 0.4571, 95% CI of partial η^2 = [0.0412 0.7403]), Orientation ($F_{1,11}$ = 14.1095, P = 0.0032, partial η^2 = 0.5619, 95% CI of partial η^2 = [0.0835 0.8458]) and interaction between them ($F_{1,11}$ = 19.3135, P = 0.0011, partial η^2 = 0.6371, 95% CI of partial η^2 = [0.3418 0.8191]). Subsequent post hoc analyses (see Statistics in Methods for details) showed a significant simple effect of Test for the exposed orientation ($F_{1,11}$ = 26.5228, P = 0.0003, partial η^2 = 0.7068, 95% CI of partial η^2 = [0.2738 0.8931]) in the NS group. On the other hand, in the FS group the results of the ANOVA showed no significant effect of Test ($F_{1,11}$ = 0.3659, P = 0.5575, partial η^2 = 0.0322, 95% CI of partial η^2 = [0.0000 0.2383]), Orientation ($F_{1,11}$ = 0.8083, P = 0.3879, partial η^2 = 0.0685, 95% CI of partial η^2 = [0.0000 0.4071]), or interaction between them ($F_{1,11}$ = 0.2955, P = 0.5976, partial η^2 = 0.0262, 95% CI of partial η^2 = [0.0000 0.2207]), indicating no significant learning in this group. These results indicate that VPL of a supra-threshold feature occurs as a result of exposure to NS images as task-irrelevant, in contrast to no performance improvement observed on the dominant orientation in the FS images as task-irrelevant and supra-threshold.

As discussed above, there has been the assumption that VPL should not occur as a result of exposure of any supra-threshold and task-irrelevant features, based on the observations using artificial features (Ahissar & Hochstein, 1993; Astorga et al., 2022; Huang et al., 2007; Lewis et al., 2009; Li et al., 2004; Mukai et al., 2011; Paffen et al., 2008; Schoups et al., 2001; Shiu &

Pashler, 1992; Tsushima et al., 2008). Our current findings indicate that the long-term controversy surrounding whether or not unsupervised learning occurs with human VPL would be resolved by assuming that the unsupervised learning occurs in a natural scene but not in an artificial stimulus.

Higher-order statistics contained in natural scene images play a critical role in task-irrelevant VPL

Then, an important question arises: What enables VPL of the dominant orientation in natural scene images to occur, whereas VPL of the dominant orientation in artificial images does not? To address this question, we further manipulated characteristics of images used in the subsequent experiments (Experiments 2-5).

In Experiments 2 and 3, we tested two distinctive characteristics of natural scene images that might enable the VPL. First, kurtosis and skewness are recognized as significant characteristics of natural scene images (Motoyoshi et al., 2007; Olshausen & Field, 1996; Simoncelli & Olshausen, 2001). This raises the possibility that VPL of FS images as task-irrelevant does not occur because it does not have the same kurtosis and skewness in the luminance histogram as in NS images. If true, we should predict exposures to FS images with the same kurtosis and skewness as NS images, termed kurtosis- and skewness-matched (KS) images (Fig. 2a, highlighted with purple rings), to lead to VPL of the dominant orientation in these images.

In Experiment 2, 12 new participants were trained and tested in the same way as in Experiment 1, with the following exception. During the exposure stage (see Fig. S4a for RSVP performance), the participants were exposed to KS images (KS group). The KS images consisted of similar local orientation and spatial frequency spectrums as NS images with kurtosis and skewness that were matched to the NS images as discussed above. Fig. 2b shows performance improvements in the KS group (see Figs. S4b and c for task accuracies before the averaging and subtraction). Results of two-way ANOVA on performance with Test (pre vs post) and Orientation (exposed vs unexposed) showed no significant effect of Test ($F_{1,11}$ = 0.6535, P = 0.4360, partial η^2 = 0.0561, 95% CI of partial η^2 = [0.0000 0.4466]), Orientation ($F_{1,11}$ = 0.0234, P = 0.8812, partial η^2 = 0.0012, 95% CI of partial η^2 = [0.0000 0.0159]), or interaction between them ($F_{1,11}$ = 0.0092, P = 0.9254, partial η^2 = 0.0008, 95% CI of partial η^2 = [0.0000 0.0052]), indicating no significant learning as a result of exposure to the KS images.

Second, the results of Experiment 2 indicate that it is not kurtosis or skewness that plays a role in VPL of the dominant orientation. A growing body of research suggests that what is known as higher-order statistics derived from NS is essential for perceiving and recognizing NS

(Freeman & Simoncelli, 2011). These higher-order statistics complement lower-order statistics, which include various local orientation and spatial frequency signals and luminance histogram that constitute of the NS, as well as the kurtosis and skewness as discussed above. The higher-order statistics include correlations between different local orientation signals and spatial frequency signals (Freeman & Simoncelli, 2011; Henderson et al., 2023; Movshon & Simoncelli, 2014). The Portella-Simoncelli (PS) images (Fig 3a, highlighted with blue rings; see Stimuli in Methods for details of creating the PS images) consist of both the lower-order and higher-order statistics derived from the NS used in Experiment 1 (Portilla & Simoncelli, 2000). If the higher-order statistics derived from NS images are necessary for VPL of the dominant orientation in these images, which are supra-threshold and task-irrelevant, exposure to the PS images should result in VPL, whereas VPL did not occur with the KS images in Experiment 2, as they lack these higher-order statistics signals.

In Experiment 3, the procedures were identical to Experiments 1-2, except that 12 new participants were exposed to the PS images during the exposure stage (see Fig. S5a for RSVP performance). PS images consisted of the higher-order statistics derived from the NS images as well as KS images (Portilla & Simoncelli, 2000). Fig. 3b shows performance improvements in the PS group (see Figs. S5b and c for task accuracies before the averaging and subtraction). Results of two-way ANOVA on performance with Test (pre vs post) and Orientation (exposed vs unexposed) showed significant effects of Test ($F_{1,11}$ = 7.0718, P = 0.0222, partial η^2 = 0.3913, 95% CI of partial η^2 = [0.0298 0.7107]) and interaction between Test and Orientation ($F_{1,11}$ = 8.6700, P = 0.0133, partial η^2 = 0.4408, 95% CI of partial η^2 = [0.1225 0.6047]), suggesting significant learning. The post hoc analyses indicated a significant simple effect of Test for the exposed dominant orientation ($F_{1,11}$ = 10.5579, P = 0.0077, partial η^2 = 0.4897, 95% CI of partial η^2 = [0.1926 0.7240]), demonstrating VPL as a result of exposure to the PS images.

The PS images consisted of the higher-order statistics and lower-order statistics, derived from NS images. Therefore, the above results indicate that PS images, consisting of both higher-order statistics and lower-order statistics derived from NS images, are sufficient to induce the task-irrelevant VPL. If task-irrelevant VPL of the dominant orientation occurs only with the higher-order statistics in PS images, we may as well say that the higher-order statistics alone are sufficient for the task-irrelevant VPL to occur.

If the higher-order statistics are sufficient for VPL of the dominant orientation to occur, the higher-order statistics alone must contain information related to the dominant orientation in the NS images. To test if the higher-order statistics alone generate information of the dominant orientation, for the first time, we created images termed a higher-order statistics (HS) image, only

consisting of the higher-order statistic signals derived from NS images (see Fig. 4a and Stimuli in Methods for details). In the HS image, the local orientation component of the original NS was practically taken away by randomizing these orientations (see Fig. S6 for orientation power distribution of HS images).

Thus, in Experiment 4, we tested whether VPL of the dominant orientation occurs as a result of exposures to the HS images. The procedures were identical to Experiments 1-3 except that 12 new participants were exposed to the HS images during the exposure stage (see Fig S7 for performance in the RSVP and orientation discrimination tasks). We applied a two-way ANOVA on performance with the factors of Test (pre- vs post-tests) and Orientation (exposed vs unexposed orientations; see Table S1 for details) and found significant effect of Test ($F_{1,11}$ = 10.8027, P = 0.0072, partial η^2 = 0.4955, 95% CI of partial η^2 = [0.1096 0.7385]) and interaction between Test and Orientation ($F_{1,11}$ = 16.1700, P = 0.0020, partial η^2 = 0.5951, 95% CI of partial η^2 = [0.3204 0.7699]), suggesting learning specific to the exposed orientation (Fig. 4b). The post hoc analyses revealed a significant simple effect of Test at the exposed orientation ($F_{1,11}$ = 15.9225, P = 0.0021, partial η^2 = 0.5914, 95% CI of partial η^2 = [0.2705 0.7647]), indicating VPL of the exposed orientation with the HS images. Notably, we found significant VPL for the dominant orientation in HS images, the higher-order images that did not contain components of lower-order statistics from original natural scene images (Fig. S6). This result supports the hypothesis that the higher-order statistics alone generate the dominant orientation based on which VPL of the orientation can occur.

In Experiments 1-4, five different types of images—NS, FS, KS, PS, and HS—were exposed. The characteristics of each type are shown in Table 1. The summary of the results (Table 2) indicates that exposures to images that contain the higher-order statistics from NS images (NS, PS, and HS) led to VPL of the dominant orientation of the images while exposures to images lacking these higher-order statistics (FS and KS) did not.

Then, why does exposure to images containing higher-order statistics from NS images lead to VPL? One possibility is that the dominant orientation generated from these higher-order statistics was sub-threshold and invisible. Task-irrelevant VPL of a feature in an artificial stimulus occurs only if the feature is sub-threshold (Chang et al., 2014; Frank et al., 2020; Tsushima et al., 2006, 2008). That might have been why task-irrelevant VPL of the dominant orientation in the images containing these higher-order statistics have occurred. To test this possibility, in Experiment 5 we investigated whether the dominant orientation in the HS images used for Experiment 4 was sub-threshold or supra-threshold. A new group of 12 participants were asked to report whether the dominant orientation in HS images for a test were rotated clockwise or

counterclockwise from the vertical (See Experiment 5 in Methods for details). The results (Fig. 5) show that the HS images were supra-threshold, ruling out the possibility.

Higher-order statistics contained in natural scene images are less subject to attentional suppression

As discussed above, consistent results show that exposure to supra-threshold task-irrelevant artificial stimuli, including Gabor patches, does not lead to VPL due to attentional suppression on the task-irrelevant stimuli (Ahissar & Hochstein, 1993; Astorga et al., 2022; Chang et al., 2014; Frank et al., 2020; Huang et al., 2007; Lewis et al., 2009; Li et al., 2004; Mukai et al., 2011; Paffen et al., 2008; Schoups et al., 2001; Shiu & Pashler, 1992; Tsushima et al., 2006, 2008). Meanwhile, the dominant orientation derived from the higher-order statistics from NS images may have spatially and/or temporarily differences in quality compared to that derived from the lower-order statistics in FS and KS images (Barbot et al., 2011; Dakin et al., 1999; Landy & Graham, 2004). This raises another possibility that the orientation derived from these higher-order statistics is less subject to attentional suppression than the orientation derived from the lower-order statistics, resulting in VPL for NS, PS, and HS images but not VPL for FS and KS images.

To test this hypothesis, in Experiment 6, we used an established experimental paradigm designed to measure the degree of attentional suppression on irrelevant features (Gaspelin et al., 2015; Stilwell et al., 2022). Specifically, we measured participants' performance in an orientation discrimination task under two conditions. In the focused attention condition, the participants focused their attention on the orientation of stimuli. In the diverted attention condition, attention was directed away from it. The degree of suppression was quantified by comparing the degrees in performance in the diverted attention condition to the focused attention condition. We also employed a third condition (RSVP condition; see below for details) for a control purpose.

A new group of 12 participants underwent the three conditions: the focused attention, diverted attention, and RSVP conditions (Fig. 6a; see Experiment 6 in Methods for details). In the focused attention condition, the participants performed the orientation discrimination task of the dominant orientation in both KS and HS images. Since KS images exclusively contain the lower-order statistics derived from NS images and HS images exclusively contain the higher-order statistics derived from these NS images, utilizing these two image types should be sufficient for comparing the degrees of suppression between task-irrelevant orientation derived from both the lower-order and higher-order statistics. In the diverted attention condition, they were asked to perform the RSVP task for 80% of the whole trials and orientation discrimination task for the remaining 20% of trials in a random order. At the beginning of each trial, participants were not

informed of which task would be assigned. The task for each trial was only informed after the stimulus offset (see Experiment 6 in Methods for details). We instructed the participants to prioritize the RSVP task over the orientation discrimination task, assuring that they diverted their attention away from the orientation at the time of the stimulus presentation. In the RSVP condition, the participants were asked to perform the RSVP task only. This RSVP condition was aimed at testing if participants indeed prioritize the RSVP task in the diverted attention condition.

To compare the degrees of suppression between the KS and HS images, we calculated behavioral suppression indices separately for the KS and HS images based on performance of the orientation discrimination task in the focused and diverted attention conditions. This index represents the percentage reduction in performance on the orientation discrimination task under the diverted attention condition, compared to the focused attention condition (see Methods for detailed calculation). Thus, a smaller value of the index indicates less suppression.

Fig. 6b shows that the behavioral suppression index for HS images is significantly smaller than that for KS images (see Fig. S8a for accuracies of the orientation discrimination task in the focused and diverted attention conditions). This finding supports the hypothesis that the supra-threshold dominant orientation in HS images is less suppressed compared to that in KS images. Notably, KS images comprise the lower-order statistics derived from NS images, while HS images comprise the higher-order statistics derived from the NS images. Furthermore, the results of Experiments 1-4 did not indicate any interaction between the lower- and higher-order statistics (see Table 2). Based on these findings, the results of Experiment 6 suggest that attentional suppression should be weaker on NS, PS, and HS, which include the higher-order statistics, than on FS and KS, which consists of the lower-order statistics.

One may raise the possibility that this lesser attentional suppression of the HS images was obtained because participants managed to increase performance in the orientation discrimination task by deprioritizing the RSVP task in the diverted attention condition specifically when the HS images were presented on the background. If so, RSVP performance in the diverted attention condition with the HS images should be significantly lower than that in other conditions. However, we found no significant difference in the RSVP performance across conditions (diverted attention vs RSVP) and background images (HS vs KS) (Fig. S8b). Thus, the lesser attentional suppression of the HS images cannot be attributed to RSVP performance.

Monkey studies have shown that the higher-order statistics derived from NS images is processed at V2 and higher areas, rather than in V1 (Freeman et al., 2013; Hatanaka et al., 2022; Okazawa et al., 2015). If neural activation representing the less suppression of the higher-order statistics indeed occurs as the above hypothesis indicates, the suppression should be observed

in the areas corresponding to those observed in the monkey studies. In Experiment 7, we tested this hypothesis.

We used fMRI to measure blood-oxygen-level-dependent (BOLD) signals in the visual areas such as V1, V2, VP, V3, V4v and V3A while a new group of 24 participants was exposed to PS and KS images (Fig. 7a; see Methods for details). During this measurement, participants were engaged in easy and harder RSVP tasks (Fig. 7a). It has been established that BOLD signals in response to task-irrelevant visual stimuli are more strongly suppressed during the harder RSVP task than the easy RSVP task (Kastner et al., 1998; O'Connor et al., 2002). Thus, we employed the easy and harder RSVP tasks to compare the degrees of attentional suppression of BOLD responses to the task-irrelevant PS and KS images. Specifically, we tested if the degrees of attentional suppression of BOLD responses differ between PS and KS images in each visual area. Since higher-order statistics derived from the NS images are the only difference between the PS and KS images (see Table 1), differences in the degrees of attentional suppression of BOLD signals between the PS and KS images should reflect processing related to suppression on the higher-order statistics in the brain.

In each visual area, which was retinotopically defined (see Methods for details), we tested whether the less suppression of the higher-order statistics is reflected in BOLD signals. For this purpose, we calculated the fMRI suppression index separately for each ROI and each of PS and KS images (Fig. 7b). This index was defined as the percentage reduction of the BOLD signal amplitudes under the harder RSVP task, compared to the easy RSVP task (see Methods for detailed calculation). A larger value of the index represents a larger degree of attentional suppression of BOLD signals. As shown in Fig. 7b (also see Fig. S9a for BOLD signal amplitudes before the calculation of the fMRI suppression indices), the areas exhibiting a smaller amount of suppression on task-irrelevant PS images compared to KS images match the areas in monkeys that responded to the higher-order statistics derived from NS images (Freeman et al., 2013; Hatanaka et al., 2022; Okazawa et al., 2015). Note that this smaller attentional suppression on the PS images cannot be attributed to performance on the RSVP task, since no significant difference in RSVP performance was observed between the PS and KS images (Fig. S9b).

In the last experiment (Experiment 8), we addressed the question of why less attentional suppression occur on a feature generated from higher-order statistics compared to a feature from lower-order statistics. One possible explanation may be that higher-order statistic signals are processed outside the temporally most effective range of attentional suppression, as illustrated in Fig. 8a. To test this possibility, we psychophysically measured the time to detect the dominant orientations in HS images and KS images (Fig. 8b; see Experiment 8 in Methods). This

measurement was performed based on a deadline paradigm where participants were asked to report their perceived orientation by a specific point in time ('deadline') for each trial (Grubb et al., 2015; Sutter & Graham, 1995; Wickelgren, 1977). We measured performance in the orientation discrimination task at various deadlines and calculated the time required for performance to reach a certain level separately for HS and KS images (see Methods for detailed calculation). We estimate that it took approximately 250 ms longer for orientations in HS images to be detected compared to for those in KS images (Fig. 8c). This result is in accordance with hypothesis that the generation of the dominant orientation representation occurs later than the time window during which attentional suppression is the strongest, although the result does not rule out other possibilities.

These results collectively suggest that exposure to NS, PS, and HS images, containing the higher-order statistics derived from NS images, leads to VPL of the dominant orientation of these images. In contrast, task-irrelevant VPL does not occur with images lacking these higher-order statistics. Weaker attentional suppression on images containing these higher-order statistics, compared to those without, may contribute to the task-irrelevant VPL of the dominant orientation in NS, PS, and HS images.

Discussion

Unsupervised learning is one the most prevailing computational models of learning, alongside supervised learning and reinforcement learning models (von der Malsburg, 1973, 1994, 1999, 2018, 2021; Willshaw et al., 1997). Whether unsupervised learning is used for human VPL has been empirically tested (Herzog & Fahle, 1998; Shibata et al., 2009, 2014; Tsodyks & Gilbert, 2004; Watanabe & Sasaki, 2015). As a result, these studies concluded that unsupervised learning model appears to be contradicted by a biological model; this conclusion is based on a significant body of findings showing that VPL does not occur as a result of exposure to supra-threshold and task-irrelevant features due to attentional suppression on those features (Astorga et al., 2022; Chang et al., 2014; Frank et al., 2020; Tsushima et al., 2008). In the current study, we demonstrate a resolution to this controversy with a new finding: Exposure to a dominant global orientation as task-irrelevant leads to VPL of the orientation, only when the orientation is derived from natural scene images. Further examinations suggest that VPL of orientation occurs if the orientation is derived from the higher-order statistics of NS images rather than from the lower-order statistics. Subsequent psychophysical and fMRI experiments showed that the orientation from the higher-order statistics is subject to a significantly smaller degree of attentional suppression than that from the lower-order statistics, particularly in areas beyond V1.

It is notable that higher-order statistics are also derived from FS and KS images. However, these higher-order statistics are far less spatially structured than those derived from NS images. This weaker structure in FS and KS is not likely to lead to the generation of the dominant orientation. In contrast, the higher-order statistics derived from NS have distinctive and meaningful structures of correlations between different orientation channels and between different spatial frequency channels at various locations, enabling them to generate the dominant orientation signal. In contrast to orientation information in lower-order statistics in FS and KS images, that from higher-order statistics in PS, HS and NS may survive attentional suppression, leading to VPL of the orientation.

Why does less attentional suppression occur on a feature generated from higher-order statistics compared to a feature from lower-order statistics? The results of Experiment 8 are in accordance with the hypothesis that higher-order statistic signals are processed outside the temporally most effective range of attentional suppression (Fig. 8b), although they do not rule out other possibilities. For instance, it is possible that attentional suppression is stronger on the lower-order statistics than on the higher-order statistics. It has been shown that first-order motion, defined by changes in luminance, tends to be more readily suppressed by attention compared to second-order motion, which is defined by changes in texture or contrast. This suggests that simpler, lower-order stimuli are more susceptible to attentional modulation than more complex, higher-order stimuli (Smith et al., 1998). Future research, which directly measure signals contributing to attentional suppression, is necessary to test these possibilities.

In our study, we investigated why VPL, as biological learning, of task-irrelevant and supra-threshold feature, does not occur, despite the prediction by unsupervised learning, a prevailing computational model, that it should occur. We found that VPL of the dominant orientation from NS images occurs, whereas VPL of an artificial stimulus, conventionally used in VPL research, does not occur. Furthermore, we found that the higher-order statistics derived from NS images enable VPL of the orientation from these statistics to occur. This is likely due to weaker attentional suppression on the orientation derived from the higher-order statistics compared to that from the lower-order statistics. These results further indicate that mechanisms of VPL for NS images are not completely the same as those of VPL of artificial stimuli, which have been used in most of VPL studies.

Figures and Tables

Table 1: Image characteristics of 5 different image types in Experiments 1-4

	Experiment 1		Experiment 2	Experiment 3	Experiment 4
Image characteristics	**Natural scene (NS)**	**Fourier-scrambled (FS)**	**Kurtosis- & skewness-matched (KS)**	**Portella-Simoncelli (PS)**	**Higher-order statistics (HS)**
Orientation	✓	✓	✓	✓	
Spatial frequency	✓	✓	✓	✓	
Kurtosis & skewness	✓		✓	✓	
Higher-order statistics	✓			✓	✓

Table 2: Summary results of Experiments 1-4

	Experiment 1		Experiment 2	Experiment 3	Experiment 4
Orientation	**Natural scene (NS)**	**Fourier-scrambled (FS)**	**Kurtosis- & skewness-matched (KS)**	**Portella-Simoncelli (PS)**	**Higher-order statistics (HS)**
Exposed orientation	Learning	No learning	No learning	Learning	Learning
Unexposed orientation	No learning	No learning	No learning	No learning	No learning

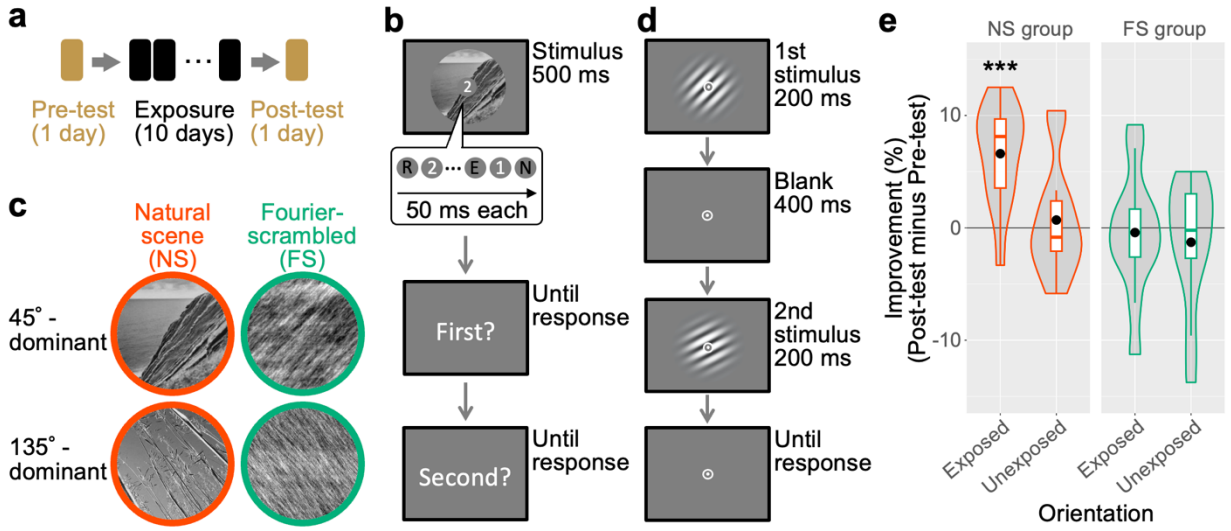

Fig. 1: Methods and results of Experiment 1

a, The complete experiment consisted of a 10-day exposure stage, preceded and followed by 1-day pre- and post-test stages. **b**, Procedure of a single trial with the rapid serial visual presentation (RSVP) task during the exposure stage. Participants were asked to report only the two white digits in a sequence of white digits and black alphabetical letters. **c**, Examples stimuli of natural scene (NS, highlighted with orange rings) and Fourier-scrambled (FS, highlighted with green rings) images presented during the exposure stage. **d**, Procedure of the orientation discrimination task in the pre-test and post-test stages. Participants were asked to report whether the orientations of the first and second stimuli were the same or different. **e**, Improvement of performance on the orientation discrimination task (post-test stage minus pre-test stage) for the NS (left panel) and FS (right panel) groups. For the NS group, results of a two-way ANOVA showed significant effect of interaction between Test (pre- vs post-tests) and Orientation (exposed vs unexposed orientations) (See Table S1 for detailed results of ANOVA). Subsequent post hoc analyses revealed a significant simple effect of Test at the exposed orientation ($F_{1,11}$ = 26.5228, P = 0.0003, partial η^2 = 0.7068, 95% CI of partial η^2 = [0.2738 0.8931]), but not at the unexposed orientation ($F_{1,11}$ = 0.2262, P = 0.6436, partial η^2 = 0.0202, 95% CI of partial η^2 = [0.0000 0.1837]). For the FS group, results of the two-way ANOVA showed no significant effect. The box plots are overlaid on the violin plots. The black dots on the box plots represents the mean accuracy across participants. The violin plots illustrate kernel probability densities of accuracies obtained from individual participants. ***P < 0.005.

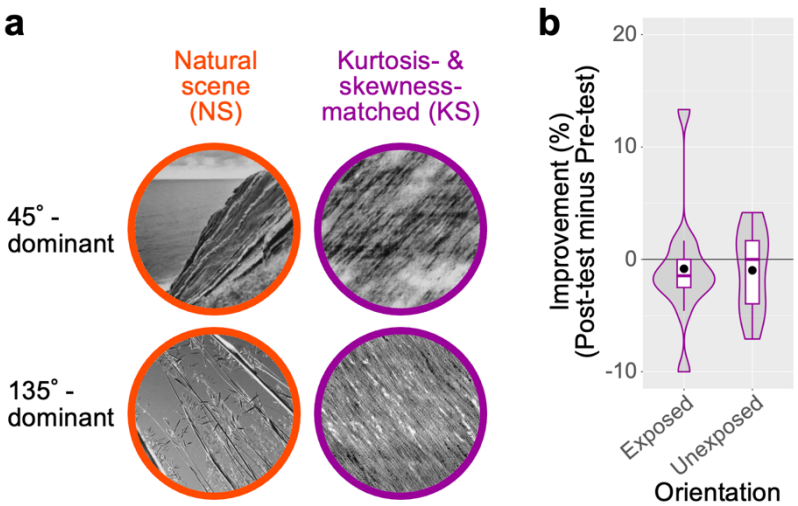

Fig. 2: Methods and results of Experiments 2

a, Examples of the NS (highlighted with orange rings) and kurtosis- & skewness-matched (KS, highlighted with purple rings) images. **b**, Improvement of performance on the orientation discrimination task (post-test stage minus pre-test stage) for the KS group. Results of the two-way ANOVA showed no significant effect of Test (pre- vs post-tests), Orientation (exposed vs unexposed orientations), or interaction between them (See Table S1 for detailed results of ANOVA). The box plots are overlaid on the violin plots. The black dots on the box plots represents the mean accuracy across participants. The violin plots illustrate kernel probability densities of accuracies obtained from individual participants.

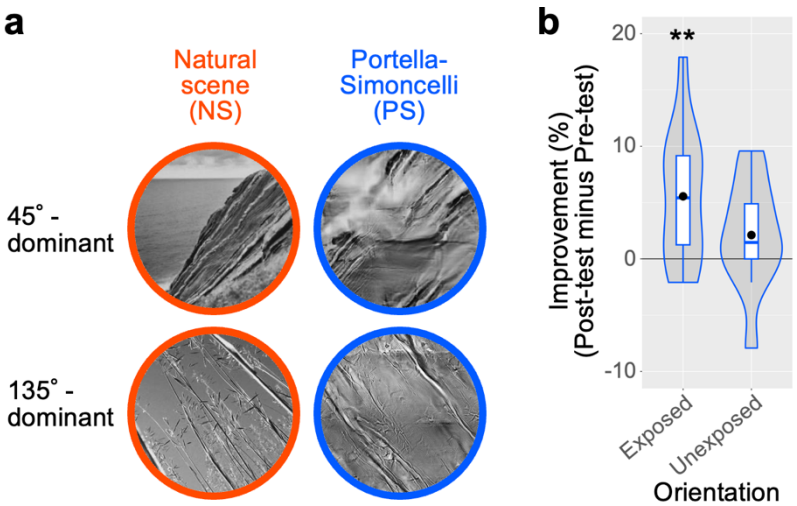

Fig. 3: Methods and results of Experiment 3

a, Examples of the NS (highlighted with orange rings) and Portilla-Simoncelli (PS, highlighted with blue rings) images. **b**, Improvement of performance on the orientation discrimination task (post-test stage minus pre-test stage) for the PS group. Results of a two-way ANOVA showed significant effect of interaction between Test (pre- vs post-tests) and Orientation (exposed vs unexposed orientations) (See Table S1 for details). Subsequent post hoc analyses revealed a significant simple effect of Test at the exposed orientation ($F_{1,11}$ = 10.5579, P = 0.0077, partial η^2 = 0.4897, 95% CI of partial η^2 = [0.1926 0.7240]), but not at the unexposed orientation ($F_{1,11}$ = 2.3348, P = 0.1547, partial η^2 = 0.1751, 95% CI of partial η^2 = [0.0000 0.5633]).The box plots are overlaid on the violin plots. The box plots are overlaid on the violin plots. The black dots on the box plots represents the mean accuracy across participants. The violin plots illustrate kernel probability densities of accuracies obtained from individual participants. **P < 0.01.

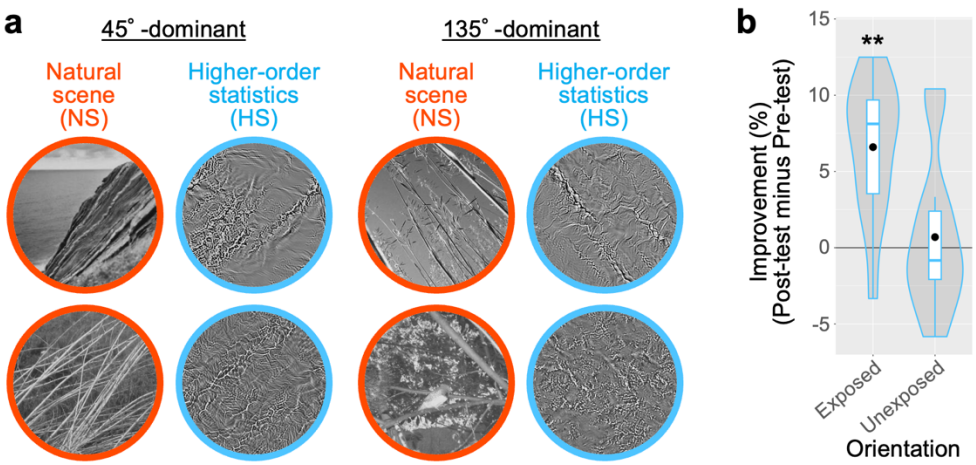

Fig. 4: Methods and results of Experiment 4

a, Examples of the NS (highlighted with orange rings) and higher-order statistics (HS; highlighted with cyan rings) images. The HS images were tailored to match the higher-order statistics of the NS images (see Stimuli in Methods for details). We created the HS images by the following steps. First, higher-order statistics of NS images were computed. Second, pixels of the NS images were randomly shuffled in a pixel space. Third, the lower-order statistics were computed using the shuffled images. Finally, the HS image was synthesized based on the computed higher-order and lower-order statistics. **b**, Improvement of performance on the orientation discrimination task (post-test stage minus pre-test stage) for the HS group. Results of a two-way ANOVA showed significant effect of Test (pre- vs post-tests) and interaction between Test and Orientation (exposed vs unexposed) (See Table S1 for details). Post hoc analyses revealed a significant simple effect of Test at the exposed orientation ($F_{1,11}$ = 15.9225, P = 0.0021, partial η^2 = 0.5914, 95% CI of partial η^2 = [0.2705 0.7647]), but not at the unexposed orientation ($F_{1,11}$ = 2.6466, P = 0.1320, partial η^2 = 0.1939, 95% CI of partial η^2 = [0.0003 0.5629]). The box plots are overlaid on the violin plots. The Black dots on the box plots indicate the mean data across participants. The violin plots illustrate kernel probability densities of data obtained from individual participants. **P < 0.01.

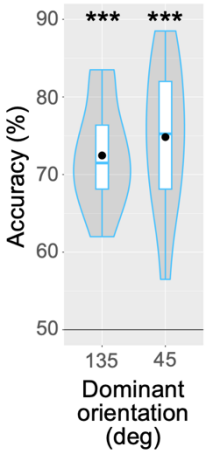

Fig. 5: Results of Experiment 5

Accuracy of the orientation discrimination task using the HS images. In each trial, participants were asked to report whether the dominant orientation of a presented image was rotated in a clockwise or counterclockwise direction from the vertical axis. Mean accuracy was significantly higher than the chance level for the 45°-dominant HS images (one-sample *t*-test with Bonferroni correction for multiple comparisons; $t_{11} = 9.2317$, corrected $P < 10^{-4}$, Cohen's d = 8.0307, 95% CI of accuracy = [68.9127 80.7540] %) and the 135°-dominant HS images ($t_{11} = 11.6124$, corrected $P < 10^{-4}$, Cohen's d = 10.8153, 95% CI of accuracy = [68.2016 76.7151] %). No significant difference was found between the orientations (paired *t*-test; $t_{11} = 0.7901$, $P < 0.4461$, Cohen's d = 0.2281, 95% CI of the difference in accuracies = [-4.2406 8.9910] %). The box plots are overlaid on the violin plots. The Black dots on the box plots indicate the mean data across participants. The violin plots illustrate kernel probability densities of data obtained from individual participants. ***$P < 0.005$.

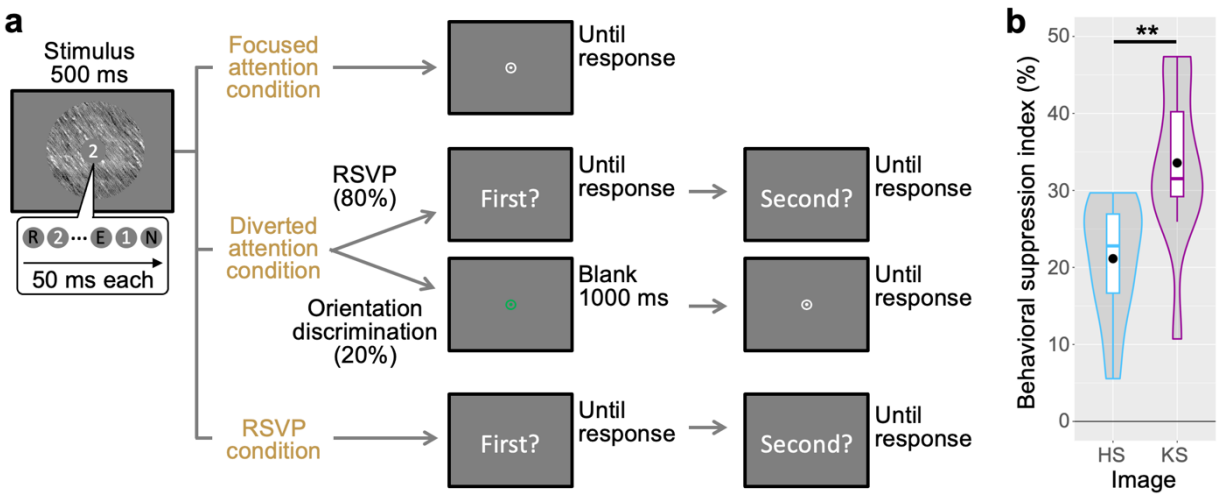

Fig. 6: Procedures and results of Experiment 6

a, Illustrative timeline of trials in the focused attention, diverted attention, and RSVP conditions. In the focused attention condition, participants were asked to report whether the dominant orientation of a presented images was rotated in a clockwise or counterclockwise direction from the vertical axis (orientation discrimination task). In the diverted attention condition, 80% of trials were assigned to the RSVP task in which participants were asked to indicate two digits in a sequence of characters. In the remaining 20% of trials, they were asked to perform orientation discrimination task. Note that the task for each trial was only informed after the stimulus offset. The participants were instructed to prioritize the RSVP task over the orientation discrimination task. In the RSVP condition, they were asked to perform the RSVP task. **b**, Behavioral suppression index computed from accuracies of the orientation discrimination task in the focused and diverted attention conditions (see Methods for detailed calculation). The mean behavioral suppression index for the HS images was significantly lower than that for the KS images (paired *t*-test; $t_{11} = 3.3072$, P = 0.0070, Cohen's d = 0.9547, 95% CI of the difference between the indices = [4.1477 20.7015] %). The box plots are overlaid on the violin plots. The black dots on the box plots indicate the mean behavioral suppression index across participants. The violin plots illustrate kernel probability densities of behavioral suppression indices obtained from individual participants. **P < 0.01.

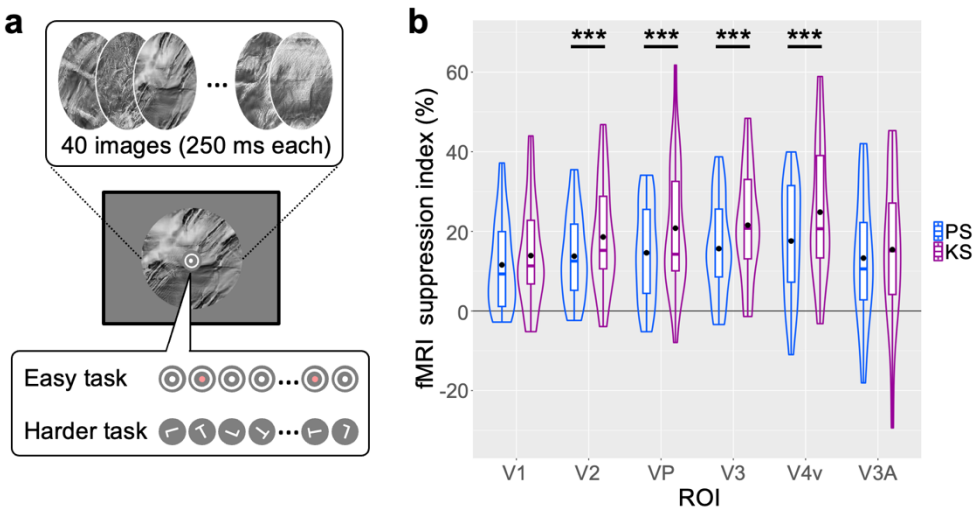

Fig. 7: Procedures and results of Experiment 7

a, Illustrative timeline of the stimulus period during which participants engaged in the easy and harder RSVP tasks while being exposed to task-irrelevant PS and KS images (see Methods for details). In the easy RSVP task, participants were required to indicate whether the number of the presentations of a red fixation point was even or odd. In the harder RSVP task, they were asked to report whether the number of the presentations of a "T" was even or odd. **b**, The fMRI suppression indices for the PS and KS images in each of regions of interest (ROIs). The fMRI suppression index was defined as the percentage reduction in activation amplitudes in the harder task, compared to the easy task (see Method for detailed calculation). A larger value of the index denotes stronger attentional suppression of a BOLD response to a task-irrelevant visual stimulus. We applied two-way ANOVA on the fMRI suppression indices with the factors of ROI (V1, V2, VP, V3, V4v, vs V3A) and Image (PS vs KS). Results of ANOVA showed significant effect of ROI ($F_{2.07,47.53}$ = 5.9282, P = 0.0046, partial η^2 = 0.2049, 95% CI of partial η^2 = [0.0918 0.3108]), Image ($F_{1,23}$ = 15.6318, P = 0.0006, partial η^2 = 0.4046, 95% CI of partial η^2 = [0.1453 0.5834]), and interaction between them ($F_{2.62,60.32}$ = 6.0818, P = 0.0017, partial η^2 = 0.2091, 95% CI of partial η^2 = [0.1072 0.3181]). Subsequent post hoc analyses revealed a significant simple effect of Image at V2 ($F_{1,23}$ = 17.8225, P = 0.0003, partial η^2 = 0.4366, 95% CI of partial η^2 = [0.1998 0.6050]), VP ($F_{1,23}$ = 16.9052, P = 0.0004, partial η^2 = 0.4236, 95% CI of partial η^2 = [0.1706 0.6024]), V3 ($F_{1,23}$ = 23.8248, P = 0.0001, partial η^2 = 0.5088, 95% CI of partial η^2 = [0.2874 0.6647]), and V4v ($F_{1,23}$ = 19.9159, P = 0.0002, partial η^2 = 0.4641, 95% CI of partial η^2 = [0.2049 0.6512]), but not at V1 ($F_{1,23}$ = 3.7372, P = 0.0656, partial η^2 = 0.1398, 95% CI of partial η^2 = [0.0006 0.3770]) or V3A ($F_{1,23}$ = 1.3660, P = 0.2511, partial η^2 = 0.0568, 95% CI of partial η^2 = [0 0.3095]). The box plots are overlaid on the violin plots. The black dots on the box plots indicate the mean data across

participants. The violin plots illustrate kernel probability densities of data obtained from individual participants. ***$P < 0.0005$.

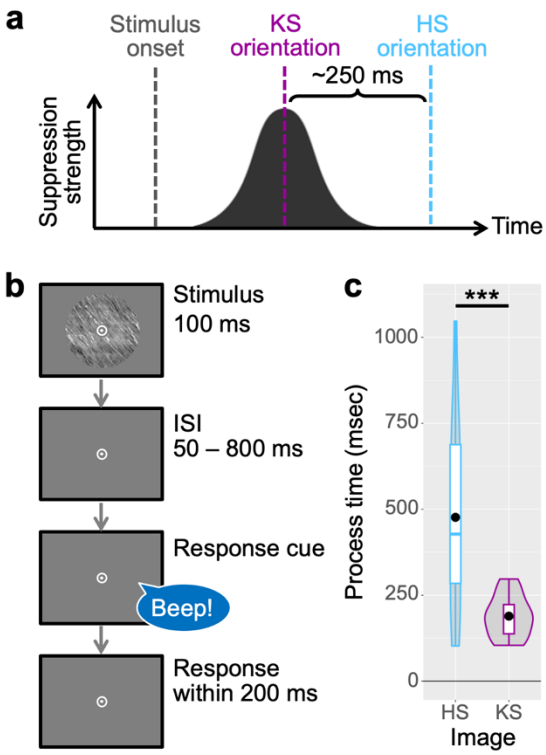

Fig 8: Procedures and results of Experiment 8

a, Graphical explanation on how suppressive signals interact with orientation information derives from the KS images ("KS orientation", magenta) and HS images ("HS orientation", cyan). Results of the psychophysical experiment showed that it took approximately 250 ms longer for the HS images to be detected compared to for the KS images (Fig. 8c). **b**, Illustrative timeline of a single trial in the orientation discrimination task, incorporating a response deadline. Participants were asked to report whether the dominant orientation of a presented image was tilted in a clockwise or counterclockwise direction with respect to the vertical axis within 200 ms after an auditory response cue. **c**, Process time computed from accuracies of the orientation discrimination task at the different inter-stimulus intervals (ISIs; see Methods for detailed computation). The mean process time for the HS images was significantly longer than that for the KS images (paired *t*-test; $t_{11} = 3.6352$, $P = 0.0039$, Cohen's d = 1.0494, 95% CI of the difference in process time = [113.3561 461.2772] ms). The box plots are overlaid on the violin plots. The Black dots on the box plots indicate the mean process time across participants. The violin plots illustrate kernel probability densities of process times obtained from individual participants. ***$P < 0.005$.

Methods

Participants

A total of 145 naïve participants with normal or corrected-to-normal vision were enrolled in this study. The experimental protocols were approved by the Institutional Review Board. Prior to their participation, all participants provided written informed consent. Data from 13 participants were subsequently omitted from the analyses (see Exclusion of participants for details). Consequently, the analyses incorporated data from 132 participants (18 to 41 years old, 82 males and 50 females).

In Experiments 1-4, we predetermined the sample size to be 12 participants per group, aligning with prior VPL studies that employed analogous designs and stimuli, as conducted by our group and other groups (Gutnisky et al., 2009; Nishina et al., 2007; Shibata et al., 2011). The sample size for the fMRI experiment (Experiment 7) was set to 24 participants, which was selected to be consistent with that used in previous fMRI studies that examined attentional modulation using ROI-based analyses (Chen et al., 2014; Gouws et al., 2014; Hickey & Peelen, 2017; Kastner et al., 2004; Mo et al., 2018).

Stimuli

Visual stimuli were presented in a monochrome scale within a visual annulus of 0.75° to 6.5° extending from the center of a uniform gray display (luminance = 0.50).

We employed oriented Gabor patches (Fig. 1d; spatial frequency = 1 cycle/degree, contrast = 100%, Gaussian filter sigma = 3°, random spatial phase) to assess participants' sensitivity to orientation. The chosen spatial frequency closely approximated the mean normalized spatial frequency observed in the natural scene images used in this study (~1 cycle/degree).

Regarding the natural scene (NS) images (Fig. 1c, highlighted with orange rings), we curated two sets of 40 images from multiple databases (Deng et al., 2009; Fei-Fei & Perona, 2005; Oliva & Torralba, 2001; Olmos & Kingdom, 2004), with each set exhibiting a dominant orientation power peaking at 45° and 135°, referred to as 45°-dominant and 135°-dominant NS images, respectively (Fig. S2). Orientation power was computed using a Sobel filter (Coppola et al., 1998).

Fourier-scrambled (FS) images (Fig. 1c, highlighted with green rings) were synthesized to replicate the orientation and spatial frequency distributions of the NS images, as well as the features derived from luminance histograms of the NS images, such as the mean and variance values. This was achieved by Fourier transforming the NS images, randomizing the phase spectra, and then applying an inverse Fourier transformation (Malach et al., 1995).

Kurtosis- and skewness-matched (KS) images (Fig. 2a, highlighted with purple rings) were generated as random images that preserved the following statistics of the NS images: orientation and spatial frequency distributions and the mean, variance, kurtosis, and skewness of luminance histograms. These images were synthesized using a pyramid-based synthesis algorithm (Heeger & Bergen, 1995) (# orientation = 4, # spatial scale = 5, window size = 15-pixel square).

The Portella-Simoncelli (PS) images (Fig. 3a, highlighted with blue rings) were generated to obtain the higher-order statistics (the correlation among the outputs of filters tuned to different positions, orientations, and scales) of the NS images, in addition to the statistics that KS images have. The PS images were synthesized using the Portella-Simoncelli algorithm (Portilla & Simoncelli, 2000) (# orientation = 4, # spatial scale = 5, window size = 15-pixel square).

The HS images (Fig. 4a, highlighted with cyan rings) were tailored to match the higher-order statistics of the NS images but differed in terms of other statistics. These images were created via the following steps. First, higher-order statistics of the NS images were computed. Second, pixels in the NS images were randomly shuffled in the pixel space. Third, other statistics (orientation and spatial frequency distributions and the mean, variance, kurtosis, and skewness of luminance histograms) were computed using the shuffled images. Finally, each HS image was synthesized based on the computed higher-order statistics and other statistics via the Portella-Simoncelli algorithm. Although derived from the 45°-dominant and 135°-dominant NS images, the HS images did not display a prominent orientation power at 45° or 135° (Fig. S6).

Experiment 1

The objective of Experiment 1 was to test whether VPL of orientation occurs as a result of exposures to supra-threshold and task-irrelevant NS and FS images. One group of 12 participants was exposed to the NS images (NS group), while the other group of 12 participants was exposed to the FS images (FS group). In each group, half of the participants were exposed to the 45°-dominant images, and the other half were exposed to the 135°-dominant images.

Experiment 1 consisted of a 10-day exposure stage, with one-day pre- and post-test stages (Fig. 1a). The purpose of the exposure stage was to expose participants to the orientation features of the NS and FS images, as task-irrelevant stimuli (Fig. 1b). To ensure that these images remained task-irrelevant, we assigned a demanding task on the fovea to participants (Chang et al., 2014; Shibata et al., 2016; Tsushima et al., 2008; Watanabe et al., 2001). In the rapid serial visual presentation (RSVP) task, participants were required to identify two digits from a rapid sequence of ten characters presented in the center of a display (Fig. 1b). Each trial began with a 500-ms fixation period. Then, ten characters were presented within a 1°-diameter circle. Each

character was displayed for 33 ms, followed by a 17-ms blank. Of these characters, two were digits (light gray, adjusted for each participant), and the remaining eight were alphabets (dark gray, luminance = 0.24). The digits and alphabets were randomly chosen from 1, 2, and 3 and A, C, D, E, G, H, J, K, M, N, P, R, S, U, V, W, and X for each trial, respectively. After the stimulus presentation, participants were required to respond by pressing buttons to indicate the first and second digits presented (no time limit). In each trial, task-irrelevant images were presented at the periphery. Forty images were presented in a random order for each block of 40 trials. Participants completed 15 blocks of the RSVP tasks per day, with brief breaks between blocks upon a participant's request.

To ensure that the RSVP task remained sufficiently demanding during the 10-day exposure stage, we adjusted the difficulty of the RSVP task in the following two ways. First, after the pre-test stage and on the same day, the participants practiced the RSVP task, and difficulty adjustments were performed. This involved testing multiple luminance levels for the digits until the accuracy reached approximately 70%, which was selected individually for each participant. Second, if a participant recorded an accuracy of over 90% on any given day of the exposure stage, the difficulty was recalibrated by modifying the luminance level, with the aim of achieving an accuracy of approximately 70% on the subsequent days.

The aim of the pre- and post-test stages was to measure changes in participants' sensitivities to orientation before and after the exposure stage. Specifically, an orientation discrimination task (Gutnisky et al., 2009) was employed (Fig. 1d). Each trial was initiated with a 500-ms fixation period, followed by the presentation of two Gabor patches for 200 ms each. These presentations were separated by a 400-ms inter-stimulus interval. Participants were asked to report whether the orientations of the two Gabor patches were the same or different by pressing a button (no time limit). Each block consisted of 48 trials. In half the trials per block, one Gabor patch was at the reference orientation (45 or 135 degrees), and the other patch was at an orientation differing by ±Δ degrees from the reference. In the other half, both Gabor patches were at the reference orientation. The reference orientation for each block was either 45 or 135 degrees, and the orientation was alternated between blocks. The Δ value for each block was 12, 8, 5, 3, or 1 degree(s), with a decrease every two blocks. Participants completed ten blocks, with brief breaks between blocks upon a participant's request. Performance in each of the pre- and post-stages was defined as the mean accuracy averaged across the Δ values of 12, 8, 5, 3, and 1 degree. An improvement in performance (Figs. 1e, 2b, 3b, and 4b) was defined by the subtraction of the performance in the pre-test stage from that in the post-test stage.

Experiments 2-4

The objective of Experiments 2-4 was to test whether VPL of orientation occurs as a result of exposures to the supra-threshold and task-irrelevant KS, PS, and HS images, respectively. The procedures were identical to those in Experiment 1, except that KS (Experiment 2), PS (Experiment 3), or HS (Experiment 4) images were used in the exposure stage. A different group of 12 participants was employed in each experiment. Half of the 12 participants were exposed to the images created from the 45°-dominant NS images, while the other half were exposed to the images created from the 135°-dominant NS images. See Stimuli in Methods for details of how the KS, PS, and HS images were created.

Experiment 5

The objective of Experiment 5 was to examine whether the orientation information contained in the HS images was sub-threshold or supra-threshold. Specifically, 12 participants engaged in an orientation discrimination task using the HS images (Fig. 4a, highlighted with cyan rings).

In each trial, following a 500-ms fixation period, an HS image was presented for 500 ms. Participants were instructed to indicate by pressing a button whether the predominant orientation of the HS image was tilted in a clockwise or counterclockwise direction with respect to the vertical axis. Each block consisted of 80 trials. In half of the trials, the HS images were created from the 45°-dominant NS images. In the other half, the HS images were created from the 135°-dominant NS images. Each participant completed five task blocks, with brief breaks between blocks upon a participant's request.

Experiment 6

The objective of Experiment 6 was to test whether attentional suppression is less likely to occur for images containing the higher-order statistics of the NS images (i.e., HS images). Twelve participants took part in an experiment using a modified version of a distractor suppression paradigm (Fig. 6a), which is typically combined with visual search tasks (Gaspelin et al., 2015; Stilwell et al., 2022).

This paradigm involved the same stimulus presentation as that in the exposure stage of Experiments 1-4, where a sequence of characters was presented on the fovea while images were presented at the periphery. Each block consisted of 40 trials. The exception in Experiment 6 was that half of the trials in each block were based on 45°-dominant images, and the other half were based on 135°-dominant images. Using this stimulus, each participant performed tasks under three conditions (Fig. 6a; focused attention, RSVP, and diverted attention) in a given order.

In the focused attention condition, as in Experiment 5, participants were asked to indicate whether the predominant orientation of the presented image was clockwise or counterclockwise compared to vertical, ignoring the sequence of characters on the fovea. This condition was conducted over two blocks, with one block presenting the HS images and the other the KS images, in a random order for each participant. A brief break was provided between the blocks upon a participant's request.

In the RSVP condition, the participants performed the same RSVP task as in Experiments 1-4 for two blocks, with a brief break between the blocks upon a participant's request. The luminance of the digits in the RSVP task was individually set for each participant, as in Experiments 1-4. As task-irrelevant stimuli, the HS images were presented in one block and the KS images were presented in another, with the order randomized for each participant.

In the diverted attention condition, participants performed the orientation discrimination task for 20% (8 trials) of each block and the RSVP task for the remaining 80% (32 trials). At the beginning of each trial, participants were not informed of which task would be assigned. The participants were informed of the task for each trial only after the stimulus offset occurred. In trials with the orientation discrimination task, a green fixation point was presented for one second following the offset of the visual stimulus. After the color of the fixation point returned to white, the participants were asked to report the predominant orientation of the image. The RSVP trials were identical to those in the original RSVP condition. Participants were instructed to prioritize the RSVP task as a primary task. A total of eight blocks were conducted in the diverted attention condition, with the HS and KS images alternately presented in each block. The order of the image presentation was randomized for each participant.

The behavioral suppression index for the orientation discrimination task was calculated for each participant and image type based on the following formula.

$$behavioral\ suppression\ index = \left(1 - \frac{accuracy_{diverted\ attention}}{accuracy_{focused\ attention}}\right) \times 100$$

Here, $accuracy_{focused\ attention}$ and $accuracy_{diverted\ attention}$ represent accuracies of the orientation discrimination task under the focused and diverted attention conditions, respectively.

Experiment 7

The objective of Experiment 7 was to identify the areas within the visual cortex that manifest differential attentional suppression in response to the higher- vs lower-order statistics of the NS images. In this experiment, 24 participants were enrolled.

We used an established paradigm (Kastner et al., 1998; O'Connor et al., 2002) to measure the effects of attentional suppression on fMRI responses to task-irrelevant stimuli (Fig. 7a). In this paradigm, while images were presented at the periphery, participants were instructed to count the number of target stimulus presentations on the fovea under two task conditions. Under one condition (easy RSVP task), the target was a fixation point that randomly turned from white to pink for 250 ms during a 10-sec stimulus period. The number of color changes during each stimulus period was randomly determined between three and six times. In the other condition (harder RSVP task), the target was the letter 'T'. During the stimulus period, a randomly rotated 'T' or 'L' was presented every 250 ms. The number of 'T's presented during the stimulus period was randomly selected between eight and fourteen. Under both conditions, participants were asked to report whether the number of target appearances was odd or even by pressing a button during the fixation period following the stimulus period. Previous research has shown that the harder RSVP task results in greater attentional suppression in fMRI responses in the visual cortex than does the easy RSVP task (Kastner et al., 1998; O'Connor et al., 2002).

For the fMRI measurements, a block design was employed. During each session, participants were instructed to focus on a white fixation point at the center of a display. Each session consisted of a 6-sec fixation period followed by 16 repetitions of a set of a 10-sec stimulus and 10-sec fixation periods (326 sec in total). During the stimulus period, 40 images were presented for 250 ms each in a random order. The stimulus periods were classified into eight types based on the image type (PS or KS), the dominant orientation of the images (45°- or 135°-dominant), and the task (easy or harder RSVP). These eight types were repeated twice in random order during the 16 stimulus periods. During the initial 4-sec interval of the fixation period, the fixation point was colored green, and participants were asked to press a button within this interval. The fixation point was turned white for the subsequent 5-sec interval, and in the last second, it changed to a 'C' or 'T.' A 'C' indicated that the next stimulus period would be under the easy RSVP task, while a 'T' indicated the harder RSVP task. Each participant underwent eight sessions, with brief breaks between the sessions upon a participant's request.

After the main fMRI session above, retinotopic mapping (Engel et al., 1997; Fize et al., 2003; Sereno et al., 1995; Tootell et al., 1997; Tootell & Hadjikhani, 2001) was also conducted to delineate regions of interest (ROIs) within the visual cortex (V1, V2, VP, V3, V4v, and V3A). We used standard retinotopic stimuli and localizer stimuli corresponding to the visual cortical areas responsive to the images used in the study.

MRI data were acquired using a 3T MRI system (Siemens, Prisma) with a 64-channel head coil. Functional images were obtained using an echo-planar imaging (EPI) sequence with

an echo time (TE) of 30 ms, a repetition time (TR) of 1,000 ms, and a flip angle (FA) of 64°. The field of view was set at 192×192 mm^2, with a matrix configuration of 64×64, and a slice thickness of 3 mm. A total of 51 contiguous slices, aligned parallel to the AC-PC line, were obtained utilizing an acceleration factor of 2 via the GRAPPA parallel imaging technique (Griswold et al., 2002) and a multiband factor of 3. We confirmed that participants were awake by monitoring their eyes through a video camera throughout the sessions. Physiological monitoring was conducted to record participants' respiratory and cardiac signals using a pressure sensor and a pulse oximeter, respectively. For the purpose of surface reconstruction, high-resolution anatomical images of the entire brain were obtained using a 3D T1-weighted MPRAGE sequence (Mugler & Brookeman, 1990) with a resolution of 0.7 × 0.7 × 0.7 mm^3 (TR = 2,180 ms, TE = 2.95 ms, inversion time = 1,100 ms, FA = 8°).

In the first stage of fMRI analysis, functional images were preprocessed. Head motion correction was performed using the AFNI command 3dvolreg (Cox, 1996) (http://afni.nimh.nih.gov/afni). Physiological noise was removed based on the participants' respiratory and cardiac signals collected during the sessions (Glover et al., 2000). The removal of slow drift components and spike noise related to head motion, and slice timing correction were performed with the first slice as a reference using in-house software. Subsequent preprocessing steps were performed using the FreeSurfer package (https://surfer.nmr.mgh.harvard.edu/). The anatomical images were normalized to a standard space. The functional images were coregistered to the anatomical image for each participant and spatially smoothed with an isotropic Gaussian kernel of 6 mm (full-width at half-maximum).

In the second stage, we performed a standard general linear model (GLM) analysis using FreeSurfer. In the GLM analysis, we modeled the four conditions for the two types of images (PS and KS) and tasks (easy and harder RSVP), as well as the fixation period and button response, as the regressors. In each ROI (V1, V2, VP, V3, V4v, and V3A), signal time series were extracted from voxels within the ROI and then averaged. The average time series for each ROI was regressed against a time series of predicted fMRI responses created by convolving a canonical hemodynamic response function with the regressors. The resultant activation amplitudes for the four conditions were defined as the contrasts between the four conditions in the stimulus period and the fixation period as a baseline. We calculated the fMRI suppression index separately for each participant, ROI (V1, V2, VP, V3, V4v, and V3A), and image type (PS, KS). The index was defined as the percentage reduction of the BOLD signal amplitudes under the harder RSVP task compared to the easy RSVP task based on the following formula.

$$fMRI\ suppression\ index = \left(1 - \frac{amplitude_{harder}}{amplitude_{easy}}\right) \times 100$$

Here, $amplitude_{easy}$ and $amplitude_{harder}$ represent BOLD signal amplitudes under the easy and harder RSVP tasks, respectively.

Experiment 8

The objective of Experiment 8 was to test whether the time required for orientation processing based on the higher-order statistics of the NS images (i.e., HS images) differed from that based on the lower-order statistics of the NS images (i.e., KS images). For this objective, 12 participants took part in an orientation discrimination task based on a deadline paradigm (Grubb et al., 2015; Sutter & Graham, 1995; Wickelgren, 1977).

In this paradigm, a 'deadline' was defined as the allocated time interval from the stimulus offset to the execution of a button response. The measurement of accuracy for various deadlines, coupled with an analysis of the time required for accuracy to reach an asymptote, enabled the quantification of time required to process a certain stimulus.

In each trial (Fig. 8b), following a 1-sec fixation period, an image was presented for 100 ms. After a predetermined inter-stimulus interval (ISI), a 50-ms high-pitched auditory cue (640 Hz) was presented, requiring participants to report the predominant orientation of the image (clockwise or counterclockwise relative to vertical) within 200 ms. Trials in which the button response exceeded this deadline were excluded from subsequent analysis. Such deadline violations occurred in 4.47 ± 0.84% (mean ± standard error) of trials across participants. Each block consisted of 40 trials. Half of the trials in each block included 45°-dominant images, and the other half included 135°-dominant images. Each participant completed 24 task blocks, with brief breaks between the blocks upon a participant's request. Two types of images (HS images and KS images) and six ISIs (50, 100, 200, 300, 500, 800 ms) were paired and assigned to each block. Across the 24 blocks, 12 pairs of the image types and ISIs were presented twice in a random order.

In the analysis, curve fitting was performed based on accuracies at the six ISI levels using the following equation:

$$Accuracy(t) = \lambda\left(1 - e^{-\beta(t-\delta)}\right) + 50,\ for\ t > \delta,\ else\ 50,$$

where t represents an ISI. Curve fitting was performed for each image type and each participant. The processing time shown in Fig. 8c was defined as $\delta + \beta^{-1}$, corresponding to the number of milliseconds required for the accuracy to reach the 67% asymptote.

Control Experiment

The objective of Control Experiment was to test the possibility that the results of Experiments 1-3 could be explained by differences in the strength of orientation information contained in the images. To examine this possibility, 12 participants performed the same orientation discrimination task (Fig. S3a) as in Experiment 5 with the NS, FS, KS, and PS images. Each block consisted of 80 trials, and the same image type was used throughout each block. Each participant completed one task block for each image type. The order of the image types was randomized for each participant. Forty images were 45°-dominant, and the other 40 images were 135°-dominant. The order of the images was randomized for each block. Brief breaks were provided between the blocks upon a participant's request.

Apparatus

All experiments were conducted in a dim room, with the experimental setup controlled via Psychtoolbox 3 (Brainard, 1997) in MATLAB. In the behavioral experiments (Experiments 1-6, 8, and Control Experiment), visual stimuli were presented on an LED display (1,920×1,080 resolution). For Experiments 1-6 and Control Experiment, the displays operated at a refresh rate of 60 Hz, whereas for Experiment 8, a higher refresh rate of 240 Hz was employed. In Experiment 7, visual stimuli were projected onto a frosted glass screen (1,920×1,080 resolution) positioned above participants' heads with a refresh rate of 60 Hz.

Exclusion of participants

In this study, data from 13 participants were excluded from the analyses for the following reasons. In Experiments 1-4, four participants were unable to continue due to health-related concerns in the middle of the exposure stage. One participant in Experiment 5 who had previously participated in a different experiment on the same day was excluded to mitigate potential biases introduced by fatigue. In Experiments 5, 6, 8, and Control Experiment, data from five participants were excluded since the experiments were terminated due to program errors. In Experiment 7, two participants were excluded due to falling asleep during the fMRI sessions. Data from one participant were excluded because the MRI measurement was terminated due to an error in the MRI system.

Statistics

All the statistical tests performed in this study were two-tailed. The significance threshold was set to an alpha level of 0.05. *t*-tests and analyses of variance (ANOVAs) were performed utilizing the

anovakun function in the R statistical software package. When a three-way ANOVA showed a significant three-way interaction, we conducted a two-way ANOVA per each level of a factor (Keselman & Keselman, 1993) to identify simple interaction effects via post-hoc analysis. Because the initial three-way ANOVA indicated a three-way interaction, this approach protected against inflating the Type 1 error rate in the post hoc analyses. This is a statistically well-established and valid way to avoid type 1 and type 2 errors (Cohen, 2004; Garson, 2012; Lubin, 1961; Rosnow & Rosenthal, 1995). Effect size for *t*-tests was quantified using Cohen's d, while partial η^2 served as the measure of effect size for ANOVA. The computation of the 95% confidence intervals was based on a bias-corrected and accelerated (BCa) method over 2,000 bootstrap repetitions. For ANOVA, we applied Medoza's multi-sample sphericity test to evaluate the sphericity assumption. When necessary, we adjusted the degrees of freedom using Greenhouse-Geisser's ε to correct any violations.

Acknowledgements

We thank K. Ueno and C. Suzuki from the Support Unit for Functional Magnetic Resonance Imaging at RIKEN Center for Brain Science for their assistance with MRI measurements. This work was supported by NSF-BSF 2241417, R01EY027841, R01EY019466 (to T.W.), R01 EY031705 (to Y.S.), JSPS Kakenhi Grants 19H01041 and 20H05715, and JST Moonshot R&D Grant JPMJMS2013 (to K.S.).

Reference

Ahissar, M., & Hochstein, S. (1993). Attentional control of early perceptual learning. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(12), 5718–5722.

Astorga, G., Chen, M., Yan, Y., Altavini, T. S., Jiang, C. S., Li, W., & Gilbert, C. (2022). Adaptive processing and perceptual learning in visual cortical areas V1 and V4. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(42), e2213080119.

Barbot, A., Landy, M. S., & Carrasco, M. (2011). Exogenous attention enhances 2nd-order contrast sensitivity. *Vision Research*, *51*(9), 1086–1098.

Barlow, H. B. (1989). Unsupervised learning. *Neural Computation*, *1*(3), 295–311.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

Chang, L.-H., Shibata, K., Andersen, G. J., Sasaki, Y., & Watanabe, T. (2014). Age-related declines of stability in visual perceptual learning. *Current Biology: CB*, *24*(24), 2926–2929.

Chen, J., He, Y., Zhu, Z., Zhou, T., Peng, Y., Zhang, X., & Fang, F. (2014). Attention-dependent early cortical suppression contributes to crowding. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(32), 10465–10474.

Cohen, B. H. (2004). *Explaining Psychological Statistics*. John Wiley & Sons.

Coppola, D. M., Purves, H. R., McCoy, A. N., & Purves, D. (1998). The distribution of oriented contours in the real world. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(7), 4002–4006.

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, an International Journal*, *29*(3), 162–173.

Dakin, S. C., Williams, C. B., & Hess, R. F. (1999). The interaction of first- and second-order cues to orientation. *Vision Research*, *39*(17), 2867–2884.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255.

Dosher, B. A., & Lu, Z. L. (1999). Mechanisms of perceptual learning. *Vision Research*, *39*(19), 3197–3221.

Doya, K. (1999). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *PERGAMON Neural Networks*, *12*, 961–974.

Dragoi, V., Turcu, C. M., & Sur, M. (2001). Stability of cortical responses and the statistics of natural scenes. *Neuron*, *32*(6), 1181–1192.

Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex* , *7*(2), 181–192.

Fei-Fei, L., & Perona, P. (2005). A Bayesian hierarchical model for learning natural scene categories. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, *2*, 524–531 vol. 2.

Fize, D., Vanduffel, W., Nelissen, K., Denys, K., Chef d'Hotel, C., Faugeras, O., & Orban, G. A. (2003). The retinotopic organization of primate dorsal V4 and surrounding areas: A functional magnetic resonance imaging study in awake monkeys. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *23*(19), 7395–7406.

Frank, S. M., Bründl, S., Frank, U. I., Sasaki, Y., Greenlee, M. W., & Watanabe, T. (2020). Fundamental Differences in Visual Perceptual Learning between Children and Adults. *Current Biology: CB*. https://doi.org/10.1016/j.cub.2020.10.047

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

Freeman, J., Ziemba, C. M., Heeger, D. J., Simoncelli, E. P., & Movshon, J. A. (2013). A functional and perceptual signature of the second visual area in primates. *Nature Neuroscience*, *16*(7), 974–981.

Garson, G. D. (2012). *General Linear Models: Univariate GLM, Anova/Ancova, Repeated Measures (Statistical Associates Blue Book Series 19)*. Statistical Associates Publishers.

Gaspelin, N., Leonard, C. J., & Luck, S. J. (2015). Direct Evidence for Active Suppression of Salient-but-Irrelevant Sensory Inputs. *Psychological Science*, *26*(11), 1740–1750.

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience*, *14*(7), 926–932.

Glaser, J. I., Benjamin, A. S., Farhoodi, R., & Kording, K. P. (2019). The roles of supervised machine learning in systems neuroscience. *Progress in Neurobiology*, *175*, 126–137.

Glover, G. H., Li, T. Q., & Ress, D. (2000). Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magnetic Resonance in Medicine:*

Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine, *44*(1), 162–167.

Gouws, A. D., Alvarez, I., Watson, D. M., Uesaki, M., Rodgers, J., & Morland, A. B. (2014). On the role of suppression in spatial attention: evidence from negative BOLD in human subcortical and cortical structures. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(31), 10347–10360.

Griswold, M. A., Jakob, P. M., Heidemann, R. M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., & Haase, A. (2002). Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, *47*(6), 1202–1210.

Grubb, M. A., White, A. L., Heeger, D. J., & Carrasco, M. (2015). Interactions between voluntary and involuntary attention modulate the quality and temporal dynamics of visual processing. *Psychonomic Bulletin & Review*, *22*(2), 437–444.

Gutnisky, D. A., Hansen, B. J., Iliescu, B. F., & Dragoi, V. (2009). Attention alters visual plasticity during exposure-based learning. *Current Biology: CB*, *19*(7), 555–560.

Hatanaka, G., Inagaki, M., Takeuchi, R. F., Nishimoto, S., Ikezoe, K., & Fujita, I. (2022). Processing of visual statistics of naturalistic videos in macaque visual areas V1 and V4. *Brain Structure & Function*, *227*(4), 1385–1403.

Heeger, D. J., & Bergen, J. R. (1995). Pyramid-based texture analysis/synthesis. *Proceedings., International Conference on Image Processing*, 229–238.

Henderson, M. M., Tarr, M. J., & Wehbe, L. (2023). A Texture Statistics Encoding Model Reveals Hierarchical Feature Selectivity across Human Visual Cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *43*(22), 4144–4161.

Herzog, M. H., & Fahle, M. (1998). Modeling perceptual learning: difficulties and how they can be overcome. *Biological Cybernetics*, *78*(2), 107–117.

Hickey, C., & Peelen, M. V. (2017). Reward Selectively Modulates the Lingering Neural Representation of Recently Attended Objects in Natural Scenes. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *37*(31), 7297–7304.

Huang, X., Lu, H., Tjan, B. S., Zhou, Y., & Liu, Z. (2007). Motion perceptual learning: when only task-relevant information is learned. *Journal of Vision*, *7*(10), 14.1-10.

Kastner, S., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science*, *282*(5386), 108–111.

Kastner, S., O'Connor, D. H., Fukui, M. M., Fehd, H. M., Herwig, U., & Pinsk, M. A. (2004). Functional imaging of the human lateral geniculate nucleus and pulvinar. *Journal of Neurophysiology*, *91*(1), 438–448.

Keselman, J. C., & Keselman, H. J. (1993). Analysis of repeated measurements. In L. K. Edwards (Ed.), *Applied Analysis of Variance in Behavioral Science* (pp. 105–145). Marcel Dekker.

Krotov, D., & Hopfield, J. J. (2019). Unsupervised learning by competing hidden units. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(16), 7723–7731.

Landy, M. S., & Graham, N. (2004). Visual Perception of Texture. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (pp. 1106–1118). MIT Press.

Lewis, C. M., Baldassarre, A., Committeri, G., Romani, G. L., & Corbetta, M. (2009). Learning sculpts the spontaneous activity of the resting human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(41), 17558–17563.

Li, W. (2016). Perceptual Learning: Use-Dependent Cortical Plasticity. *Annual Review of Vision Science*, *2*, 109–130.

Li, W., Piëch, V., & Gilbert, C. D. (2004). Perceptual learning and top-down influences in primary visual cortex. *Nature Neuroscience*, *7*(6), 651–657.

Lu, Z.-L., & Dosher, B. A. (2022). Current directions in visual perceptual learning. *Nature Reviews Psychology*, *1*(11), 654–668.

Lubin, A. (1961). The Interpretation of Significant Interaction. *Educational and Psychological Measurement*, *21*(4), 807–817.

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., Ledden, P. J., Brady, T. J., Rosen, B. R., & Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(18), 8135–8139.

Matteucci, G., Piasini, E., & Zoccolan, D. (2024). Unsupervised learning of mid-level visual representations. *Current Opinion in Neurobiology*, *84*, 102834.

Mo, C., He, D., & Fang, F. (2018). Attention Priority Map of Face Images in Human Early Visual Cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *38*(1), 149–157.

Motoyoshi, I., Nishida, S., Sharan, L., & Adelson, E. H. (2007). Image statistics and the perception of surface qualities. *Nature*, *447*(7141), 206–209.

Movshon, J. A., & Simoncelli, E. P. (2014). Representation of Naturalistic Image Structure in the Primate Visual Cortex. *Cold Spring Harbor Symposia on Quantitative Biology*, *79*, 115–122.

Mugler, J. P., 3rd, & Brookeman, J. R. (1990). Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE). *Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, *15*(1), 152–157.

Mukai, I., Bahadur, K., Kesavabhotla, K., & Ungerleider, L. G. (2011). Exogenous and endogenous attention during perceptual learning differentially affect post-training target thresholds. *Journal of Vision*, *11*(1). https://doi.org/10.1167/11.1.25

Nishina, S., Seitz, A. R., Kawato, M., & Watanabe, T. (2007). Effect of spatial distance to the task stimulus on task-irrelevant perceptual learning of static Gabors. *Journal of Vision*, *7*(13), 2.1-10.

O'Connor, D. H., Fukui, M. M., Pinsk, M. A., & Kastner, S. (2002). Attention modulates responses in the human lateral geniculate nucleus. *Nature Neuroscience*, *5*(11), 1203–1209.

Okazawa, G., Tajima, S., & Komatsu, H. (2015). Image statistics underlying natural texture selectivity of neurons in macaque V4. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(4), E351-60.

Oliva, A., & Torralba, A. (2001). Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. *International Journal of Computer Vision*, *42*(3), 145–175.

Olmos, A., & Kingdom, F. A. A. (2004). A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception*, *33*(12), 1463–1473.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*(6583), 607–609.

Paffen, C. L. E., Verstraten, F. A. J., & Vidnyánszky, Z. (2008). Attention-based perceptual learning increases binocular rivalry suppression of irrelevant visual features. *Journal of Vision*, *8*(4), 25.1-11.

Portilla, J., & Simoncelli, E. P. (2000). A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients. *International Journal of Computer Vision*, *40*(1), 49–70.

Rosnow, R. L., & Rosenthal, R. (1995). "Some Things You Learn Aren't So": Cohen's Paradox, Asch's Paradigm, and the Interpretation of Interaction. *Psychological Science*, *6*(1), 3–9.

Sagi, D., & Tanne, D. (1994). Perceptual learning: learning to see. *Current Opinion in Neurobiology*, *4*(2), 195–199.

Sasaki, Y., Nanez, J. E., & Watanabe, T. (2010). Advances in visual perceptual learning and plasticity. *Nature Reviews. Neuroscience*, *11*(1), 53–60.

Schoups, A., Vogels, R., Qian, N., & Orban, G. (2001). Practising orientation identification improves orientation coding in V1 neurons. *Nature*, *412*(6846), 549–553.

Seitz, A. R., & Dinse, H. R. (2007). A common framework for perceptual learning. *Current Opinion in Neurobiology*, *17*(2), 148–153.

Seitz, A. R., & Watanabe, T. (2003). Psychophysics: Is subliminal learning really passive? *Nature*, *422*(6927), 36.

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., Rosen, B. R., & Tootell, R. B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, *268*(5212), 889–893.

Shibata, K., Sagi, D., & Watanabe, T. (2014). Two-stage model in perceptual learning: toward a unified theory. *Annals of the New York Academy of Sciences*, *1316*, 18–28.

Shibata, K., Sasaki, Y., Kawato, M., & Watanabe, T. (2016). Neuroimaging Evidence for 2 Types of Plasticity in Association with Visual Perceptual Learning. *Cerebral Cortex* , *26*(9), 3681–3689.

Shibata, K., Watanabe, T., Sasaki, Y., & Kawato, M. (2011). Perceptual learning incepted by decoded fMRI neurofeedback without stimulus presentation. *Science*, *334*(6061), 1413–1415.

Shibata, K., Yamagishi, N., Ishii, S., & Kawato, M. (2009). Boosting perceptual learning by fake feedback. *Vision Research*, *49*(21), 2574–2585.

Shiu, L. P., & Pashler, H. (1992). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Perception & Psychophysics*, *52*(5), 582–588.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, *24*, 1193–1216.

Smith, A. T., Greenlee, M. W., Singh, K. D., Kraemer, F. M., & Hennig, J. (1998). The processing of first- and second-order motion in human visual cortex assessed by functional magnetic

resonance imaging (fMRI). *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *18*(10), 3816–3830.

Stilwell, B. T., Egeth, H., & Gaspelin, N. (2022). Electrophysiological Evidence for the Suppression of Highly Salient Distractors. *Journal of Cognitive Neuroscience*, *34*(5), 787–805.

Sutter, A., & Graham, N. (1995). Investigating simple and complex mechanisms in texture segregation using the speed-accuracy tradeoff method. *Vision Research*, *35*(20), 2825–2843.

Tootell, R. B., & Hadjikhani, N. (2001). Where is “dorsal V4” in human visual cortex? Retinotopic, topographic and functional evidence. *Cerebral Cortex* , *11*(4), 298–311.

Tootell, R. B., Mendola, J. D., Hadjikhani, N. K., Ledden, P. J., Liu, A. K., Reppas, J. B., Sereno, M. I., & Dale, A. M. (1997). Functional analysis of V3A and related areas in human visual cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *17*(18), 7060–7078.

Tsodyks, M., & Gilbert, C. (2004). Neural networks and perceptual learning. *Nature*, *431*(7010), 775–781.

Tsushima, Y., Sasaki, Y., & Watanabe, T. (2006). Greater disruption due to failure of inhibitory control on an ambiguous distractor. *Science*, *314*(5806), 1786–1788.

Tsushima, Y., Seitz, A. R., & Watanabe, T. (2008). Task-irrelevant learning occurs only when the irrelevant feature is weak. *Current Biology: CB*, *18*(12), R516-7.

von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, *14*(2), 85–100.

von der Malsburg, C. (1994). The Correlation Theory of Brain Function. In E. Domany, J. L. van Hemmen, & K. Schulten (Eds.), *Models of Neural Networks: Temporal Aspects of Coding and Information Processing in Biological Systems* (pp. 95–119). Springer New York.

von der Malsburg, C. (1999). The what and why of binding: the modeler’s perspective. *Neuron*, *24*(1), 95–104, 111–125.

von der Malsburg, C. (2018). Concerning the Neural Code. In *arXiv [q-bio.NC]*. http://arxiv.org/abs/1811.01199

von der Malsburg, C. (2021). Toward understanding the neural code of the brain. *Biological Cybernetics*, *115*(5), 439–449.

Wang, X., Zhou, Y., & Liu, Z. (2013). Transfer in motion perceptual learning depends on the difficulty of the training task. *Journal of Vision*, *13*(7), 5.

Watanabe, T., Náñez, J. E., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature*, *413*(6858), 844–848.

Watanabe, T., & Sasaki, Y. (2015). Perceptual learning: toward a comprehensive theory. *Annual Review of Psychology*, *66*, 197–221.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, *41*(1), 67–85.

Willshaw, D. J., Von Der Malsburg, C., & Longuet-Higgins, H. C. (1997). How patterned neural connections can be set up by self-organization. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *194*(1117), 431–445.