

# Efficient coding of natural images in the mouse visual cortex

Received: 21 July 2022

Accepted: 6 February 2024

Published online: 19 March 2024

 Check for updates

Federico Bolaños<sup>1</sup>, Javier G. Orlandi<sup>2</sup>✉, Ryo Aoki<sup>3</sup>, Akshay V. Jagadeesh<sup>4</sup>, Justin L. Gardner<sup>4</sup> & Andrea Benucci<sup>3,5</sup>✉

How the activity of neurons gives rise to natural vision remains a matter of intense investigation. The mid-level visual areas along the ventral stream are selective to a common class of natural images—textures—but a circuit-level understanding of this selectivity and its link to perception remains unclear. We addressed these questions in mice, first showing that they can perceptually discriminate between textures and statistically simpler spectrally matched stimuli, and between texture types. Then, at the neural level, we found that the secondary visual area (LM) exhibited a higher degree of selectivity for textures compared to the primary visual area (V1). Furthermore, textures were represented in distinct neural activity subspaces whose relative distances were found to correlate with the statistical similarity of the images and the mice's ability to discriminate between them. Notably, these dependencies were more pronounced in LM, where the texture-related subspaces were smaller than in V1, resulting in superior stimulus decoding capabilities. Together, our results demonstrate texture vision in mice, finding a linking framework between stimulus statistics, neural representations, and perceptual sensitivity—a distinct hallmark of efficient coding computations.

Visual textures are broadly defined as “pictorial representations of spatial correlations”<sup>1</sup>—images of materials with orderly structures and characteristic statistical dependencies. They are pervasive in natural environments, playing a fundamental role in the perceptual segmentation of the visual scene<sup>1,2</sup>. For example, textures can emphasize boundaries, curvatures<sup>3,4</sup>, 3D tilts, slants<sup>5,6</sup> and distortions, support a rapid “pop-out” of stimulus features<sup>7</sup>, and can form a basis set of visual features necessary for object vision<sup>8</sup>.

Although texture images largely share the spectral complexity of other natural images<sup>9–11</sup>, they can be more conveniently parametrized and synthesized than other natural images. This has been explored via diverse computational approaches: in the field of computer graphics<sup>12</sup>, via entropy-based methods<sup>13–15</sup>, using wavelet approaches<sup>16,17</sup>, and, more recently, in machine learning implementations based on deep convolutional neural networks<sup>18–21</sup>.

In light of their rich statistics and convenient synthesis and parametrization, texture images have been at the core of studies on efficient coding principles of neural processing. According to one interpretation of the efficient coding hypothesis<sup>22</sup>, the processing of visual signals along hierarchically organized cortical visual areas reflects the statistical characteristics of the visual inputs that these neural circuits have learned to encode, both developmentally and evolutionarily<sup>23–29</sup>. Accordingly, texture images have been extensively used in experimental studies that have examined the contribution of different visual areas to the processing of texture statistics.

In particular, studies in primates have revealed that the “mid-level” ventral areas, V2–V4, are crucial for processing texture images<sup>30–41</sup>, more so than the primary visual cortex, V1 (however, see ref. 42). Furthermore, as revealed by psychophysical observations<sup>43</sup> and neural measurements, area V2, in addition to being differentially

<sup>1</sup>University of British Columbia, Neuroimaging and NeuroComputation Centre, Vancouver, BC V6T, Canada. <sup>2</sup>University of Calgary, Department of Physics and Astronomy, Calgary, AB T2N 1N4, Canada. <sup>3</sup>RIKEN Center for Brain Science, Laboratory for Neural Circuits and Behavior, Wakoshi, Japan. <sup>4</sup>Stanford University, Wu Tsai Neurosciences Institute, Stanford, CA, USA. <sup>5</sup>Queen Mary, University of London, School of Biological and Behavioral Science, London E1 4NS, UK. ✉e-mail: [javier.orlandi@ucalgary.ca](mailto:javier.orlandi@ucalgary.ca); [andrea@benuccilab.net](mailto:andrea@benuccilab.net)

modulated by the statistical dependencies of textures, correlates with the perceptual sensitivity for these stimuli<sup>34,35,38</sup>. Notably, biology-inspired computational studies using artificial neural networks have similarly emphasized hierarchical coding principles, with V2-like layers as the locus for representing texture images in classification tasks<sup>44,45</sup>. Together, these observations suggest a general hierarchical coding framework, where the extrastriate visual areas, in particular area V2, define a neural substrate for representing texture stimuli, reflecting a progressive elaboration of visual information from “lower” to “higher” areas along the ventral visual stream.

This high-level view raises two fundamental questions: (1) whether this coding framework applies, in all generality, to hierarchically organized visual architectures as seen in several mammalian species other than primates—as CNN simulations would suggest—and (2) which functional principles at the circuit level give rise to texture selectivity, especially in the secondary visual area V2. Both questions hinge on the need to gain a computational and mechanistic understanding of how visual networks process naturalistic statistical dependencies to enable the perception of scenes and objects<sup>1,2,46–48</sup>.

Addressing these questions in the mouse model organism would be particularly advantageous<sup>49</sup>. Although the rodent visual system is much simpler than that of primates<sup>50</sup>, mice and rats have a large secondary visual cortex (area LM) homologous to primate V2<sup>51,52</sup>, belonging to a set of lateral visual areas forming a ventral stream of visual processing<sup>53,54</sup>. As recordings from these areas have revealed, there is increased selectivity for complex stimulus statistics in both rats<sup>55,56</sup> and mice<sup>57,58</sup>. Therefore, we studied the processing of texture images in mice with an emphasis on the interrelationship between behavioral, neural, and stimulus-statistic representations. Using a CNN-based algorithm for texture synthesis<sup>59</sup>, we generated an arbitrary number of naturalistic texture exemplars and “scrambles”—spectrally matched images lacking the higher-order statistical complexity of textures<sup>48,60–63</sup>—by precisely controlling the statistical properties of all the images. Using these images, we demonstrated that mice can perceptually detect higher-order statistical dependences in textures, distinguishing them from scrambles, and discriminating among the different types of naturalistic textures (“families” hereafter). At the neural level, using mesoscopic and two-photon GCaMP imaging, we found that the area LM was differentially modulated by texture statistics, more so than V1 and other higher visual areas (HVAs). Examining the representational geometry of the population responses, we found that when the statistical properties of a texture were most similar to those of scrambles, the corresponding neural activity was also more difficult to decode, and the animal’s performance decreased. These dependencies were particularly prominent in LM and when considering the higher-order statistical properties of the images. Notably, LM encoded different texture families in neural subspaces that were more compact than in V1, thus enabling better stimulus decoding in this area.

## Results

### Training mice to detect and discriminate between texture statistics

To examine the ability of mice to use visual–texture information during perceptual behaviors, we designed two go/no go tasks. In the first task, mice had to detect the texture images interleaved with scramble stimuli. In the second task, mice had to discriminate between two types of texture images from different texture families.

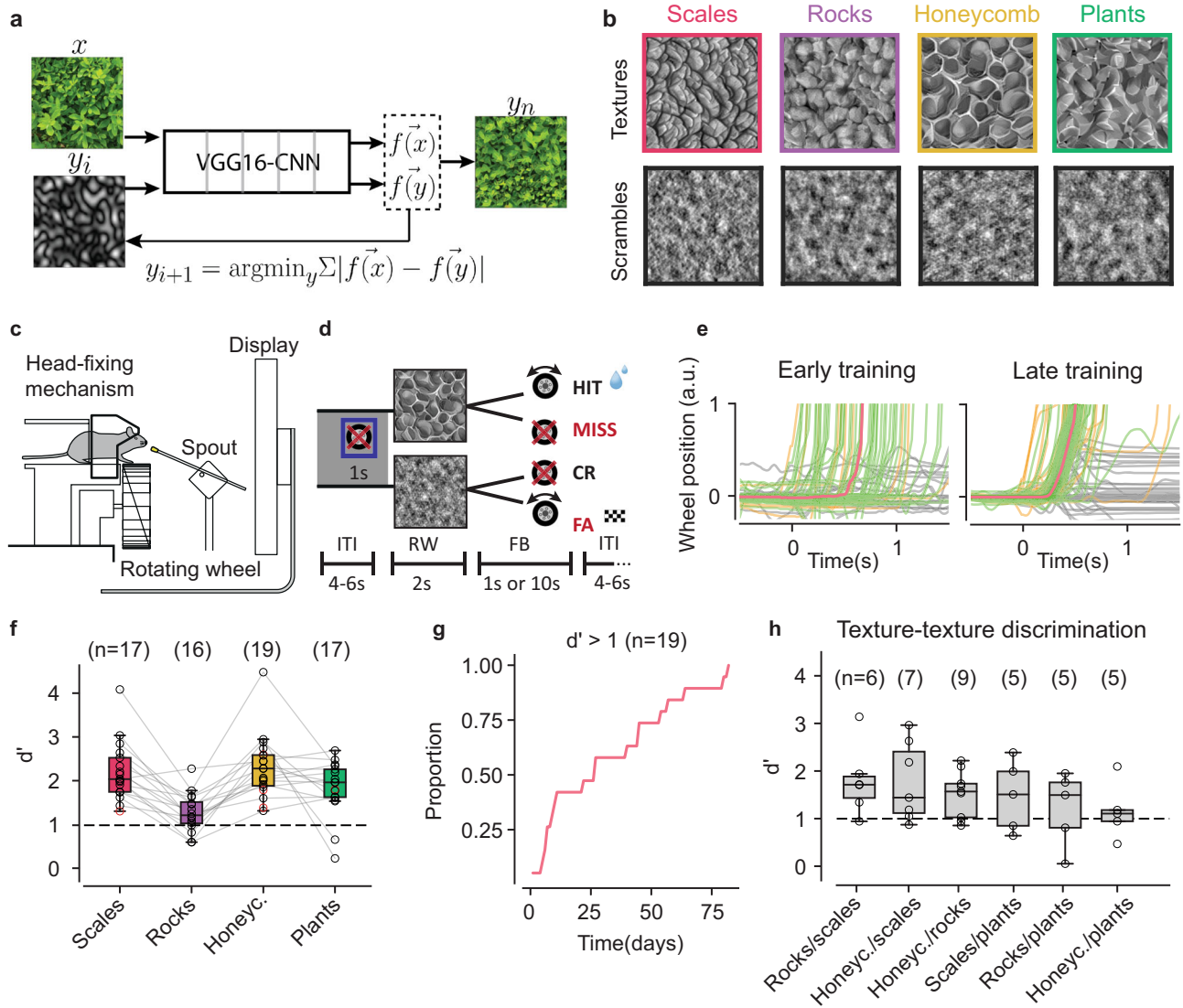
**Synthesis of textures and scrambles.** We generated synthetic textures using an iterative model that uses a convolutional neural network (VGG16) to extract a compact multi-scale representation of texture images<sup>59</sup> (Fig. 1a). To disentangle the contribution of higher-order image statistics from lower-order ones, for each texture exemplar we synthesized a spectrally matched image (scramble, Fig. 1b) having the

same mean luminance, contrast, average spatial frequency, and orientation content (Supplementary Fig. 1a–c, Methods) but lacking the higher-order statistical features characteristic of texture images. This produced image pairs for which the main axis of variation was higher-order statistics (textural information). In total, we synthesized images belonging to four texture families and four associated scramble families, each with 20 exemplars.

**Behavioral detection of texture statistics.** To train the mice in the two go/no go tasks, we employed an automated training setup<sup>64,65</sup>, wherein the mice were asked to self-head fix and respond to the visual stimuli displayed on a computer screen located in front of them (Fig. 1c). Mice were trained to respond to the target stimuli by rotating a toy wheel, and contingent on a correct response, they were rewarded with water. For the texture/scramble go/no go task, the “go” stimuli were texture images, while the “no go” stimuli were image scrambles (Fig. 1d). For responses to a no-go stimulus (false alarms), a checkerboard pattern was displayed on the screen for 10 s before a new trial began. All mice ( $n = 19$ ) learned the task in approximately 25 days (i.e., the time needed for  $d\text{-prime} > 1$  in at 50% of the mice, Fig. 1e–g). Mice could significantly discriminate between all four texture/scramble pairs (Fig. 1f,  $d' > 1$ ,  $p < 0.05$  for all families, one sample t-test with Holm-Bonferroni correction for multiple comparisons; Supplementary Table 1) with an average discriminability value of  $d' = 2.1 \pm 0.15$  (s.e.), and with the “rocks” family having a significantly lower performance than all other families, both within and across mice. Dissecting the animals’ performance, we found that, on average, mice had a high proportion of hits (Supplementary Fig. 2a), as expected given that the training procedure encouraged “go” behaviors<sup>66</sup>, with the lowest performance for rocks associated with a higher proportion of false alarms (Supplementary Fig. 2b). Additionally, to ensure that the mice were not adopting a strategy based on “brute force” memorization (e.g., of pixel-level luminance features<sup>67</sup>), we synthesized a novel image set consisting of 20 new exemplars for each of the four families, together with corresponding scramble images. Then, in a subset of the mice ( $n = 4$  for scales;  $n = 3$ , rocks;  $n = 11$ , honeycomb;  $n = 8$ , plants), we switched the original set of images with the novel set and compared behavioral performance between the last five sessions prior to the switch and the five sessions after the switch, finding no significant difference (Supplementary Fig. 2c).

**Behavioral discrimination between texture families.** Mice not only could detect higher-order statistical features in texture images that were missing in the scrambles, but they could also discriminate between different texture statistics. We trained mice already expert in texture–scramble discrimination, as well as a new cohort of naïve mice ( $n = 2$ ), in a second go/no go task (Supplementary Table 2). Mice were shown exemplars ( $n = 20$ ) from two texture families, randomly chosen but fixed across sessions, with only one of the two families associated with a water reward for a correct “go” response. In addition, all 40 exemplars were randomly rotated to prevent mice from solving this task using orientation information that may have been different across families (Supplementary Fig. 2d). Mice could discriminate between texture families, with a significantly positive  $d'$  for all six texture pairs (Fig. 1h,  $d' > 0$ ,  $p < 0.019$  for all pairs, one sample t-test with Holm-Bonferroni correction at  $\alpha = 0.05$ ).

Finally, we controlled that in both tasks mice were not relying on “simple” statistics, such as the skewness and kurtosis of the luminance histogram, with skewness having been previously related to texture perception (e.g., the *blackshot* mechanism<sup>61</sup>). For this, we created a new set of textures and scrambles in which skewness and kurtosis values were randomly mixed between textures and scrambles, thus uninformative of the texture family and texture-scramble identity (Supplementary Fig. 3a), finding that behavioral performance was unaffected by this manipulation (Supplementary Fig. 3b). This result



**Fig. 1 | Mice can discriminate texture statistics from spectrally matched scrambles and between texture families.** **a** Schematic plot of the iterative algorithm to synthesize the texture images based on the VGG16-CNN architecture; 'x', target texture;  $f(x)$ , texture representation by the network; 'y' is an initial 'seeding' Gaussian-noise image with  $f(y)$  being its network representation. The optimization minimizes the difference between  $f(x)$  and  $f(y)$  by iteratively changing 'y' to obtain 'y<sub>n</sub>'. **b** Examples of texture families and the respective spectrally matched stimuli (scrambles). **c** Schematic plot of the automatic training system with self-head fixation. **d** Texture - scramble go/no go task: the mouse must rotate a rubber wheel (go trial) if shown a texture exemplar; it must keep it still if shown a scramble (no go). ITI is the inter-trial interval; RW, the response window; and FB, the feedback period. **e** Representative examples of wheel rotations from an early training session (left) and a well-trained mouse (right); green for hits, yellow for false alarms, gray for either misses or correct rejects, and orange for the average across hits. **f** Behavioral discriminability ( $d'$ ) in the texture-scramble task for expert mice for

each family; colors as in **b**. The top labels are the number of mice trained in each of the families;  $n = 16$  out of 19 mice were trained in all the family-scramble pairs (connecting gray lines). The empty dots indicate the individual animals. Rocks have the lowest performance: one-way ANOVA across families,  $p = 2 \times 10^{-5}$ ; post-hoc Tukey HSD, rocks vs. honeycomb:  $p = 3 \times 10^{-5}$ , rocks vs. plants:  $p = 0.042$ , rocks vs. scales:  $p = 0.0002$ ,  $n = 16$  mice. Rocks,  $d' = 1.4 \pm 0.14$  tested for  $d' > 1$  (horizontal broken line),  $p = 0.016$ , one sample t-test, two-sided. **g** The time needed for mice (proportion of days) to reach  $d' > 1$  in their first family-scramble training. **h** Behavioral discriminability ( $d'$ ) in the texture-texture task across all six possible pairs of the four families. The top labels are the number of mice trained in each texture pair; broken horizontal line as in (f). Each animal was trained in a different number of family pairs (Supplementary Table 1, 2). Box plots in (f, h) and in other figures indicate the median with a horizontal bar; the box height denotes the inter-quartile range (IQR, 1<sup>st</sup> and 3<sup>rd</sup> quartile) and the whiskers extend by 1.5 x IQR. Source data are provided as a Source Data file.

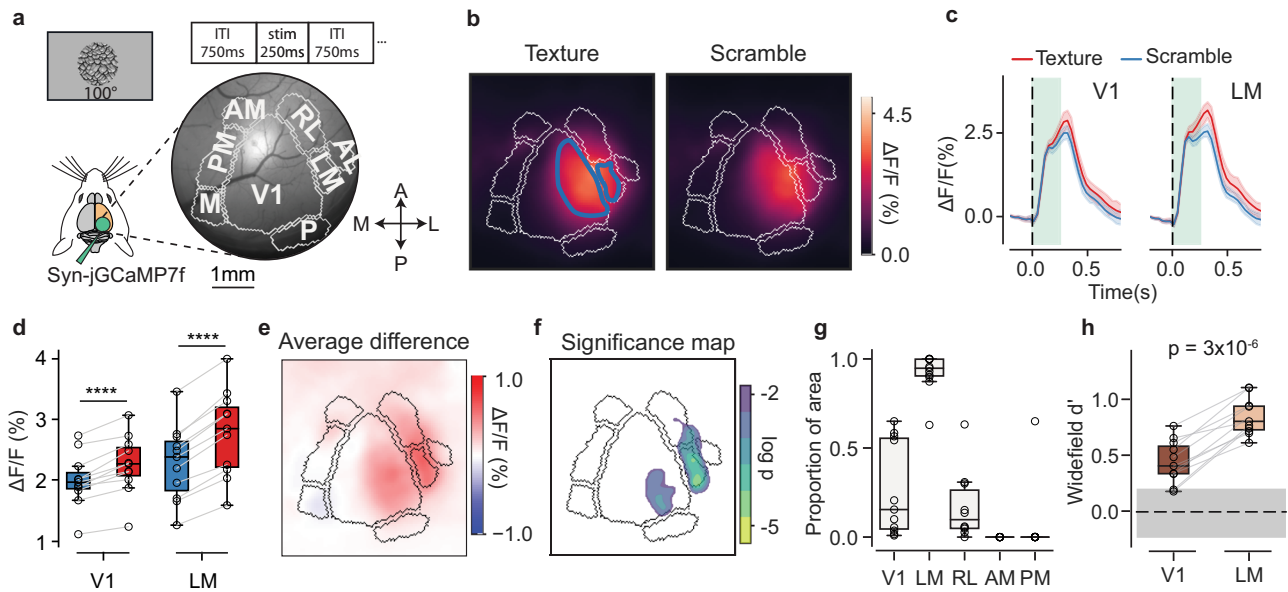
shows that the heuristic<sup>68</sup> employed by the mice was not based on these simple statistics, supporting the interpretation that mice relied on the high-order spatial correlation properties of texture images.

**Widefield responses to textures and scrambles**

To examine the neural activity underlying the mice's ability to detect and discriminate between texture statistics, we imaged multi-area responses from the posterior cortex of untrained animals whose neural dynamics were unaffected by procedural or perceptual learning processes. This choice assumes that texture processing in visual

cortical networks is likely not the outcome of our behavioral training (see also Discussion).

We performed widefield calcium imaging during the passive viewing of textures and scrambles. Mice ( $n = 11$ ) were placed in front of a computer screen that displayed either an exemplar of a texture or a scramble (Fig. 2a). The stimuli, 100 degrees in size, were presented in front of the mice, centered on the mouse's body midline, as was done for behavioral training. While mice passively viewed the stimuli, we recorded both calcium-dependent and calcium-independent GCaMP responses using a dual wavelength imaging setup. We then used the



**Fig. 2 | Texture stimuli differentially modulate V1 and LM at the mesoscale level.** **a** Schematic plot of the widefield imaging setup. Top, an example of a texture image (100° diameter) and the stimulus presentation times. Right, a representative example of the right posterior cortex of a mouse with main-area borders (gray lines). V1, primary visual cortex; LM, secondary visual cortex (lateromedial); RL, rostralateral; AL, anterolateral; AM, anteromedial; PM, posteromedial; M, medial area; P, posterior area. **b** Average GCaMP response (across all exemplars and repeats) to texture stimuli (left) and scramble stimuli (right) for a representative example mouse. The blue contours show the regions of interest (ROIs) retinotopically matching the visual stimuli in V1 and LM. **c** Average GCaMP responses from the retinotopically matched ROIs in V1 and LM for the same representative example shown in (b): texture–scramble difference, % $\Delta F/F$ . V1:  $0.27 \pm 0.04\%$ ; LM:  $0.50 \pm 0.03\%$ . Textures in red, scrambles in blue. The green-shaded rectangles show the stimulus duration (250 ms); the vertical broken line indicates the time of stimulus onset. Shaded bands are 95% CIs (4 families  $\times$  20 exemplars). **d**, Responses across mice to textures and scrambles in V1 and LM ROIs; V1: scramble,  $\Delta F/F$ :  $2.0 \pm 0.1\%$  s.e., texture:  $2.2 \pm 0.1\%$ ;  $p = 3 \times 10^{-5}$ , two-sided

paired t-test,  $n = 11$  mice. For LM: scramble:  $2.3 \pm 0.2$ , texture:  $2.8 \pm 0.2$ ;  $p = 1 \times 10^{-6}$ . Color code as in (c). Box plots indicate the median with a horizontal bar; the box height denotes the inter-quartile range (IQR, 1st and 3rd quartile) and the whiskers extend by  $1.5 \times$  IQR. **e**, The difference between the texture–scramble images shown in (b). **f**, The regions with a statistically significant response difference shown in (b). **g**, The regions with a statistically significant response difference shown in (b) – the logarithm of the p-values from a two-sided t-test; colored regions for  $p < 0.01$ . **g**, Proportion of pixels in each visual area (within the retinotopically identified ROIs) modulated by the textures relative to the scrambles ( $n = 11$  mice, empty dots); V1:  $0.27 \pm 0.08$  s.e., LM:  $0.92 \pm 0.03$ ; RL:  $0.17 \pm 0.06$ ; AM: not detected; PM:  $0.06 \pm 0.06$ . Box plots as in (d). **h** Discriminability measure ( $d'$ ) between textures and scrambles from  $\Delta F/F$  (%) responses within the same ROIs used for (g). The gray horizontal band corresponds to a null  $d'$  distribution derived from pre-stimulus activity. Each dot is one animal, connecting lines for the same-mouse data. The horizontal broken line indicates the mean of the null distribution; p-value from two-sided paired t-test. For statistical significance in this and other figures. \* is for  $p < 10^{-2}$ , \*\*  $p < 10^{-3}$ , \*\*\*  $p < 10^{-4}$ , \*\*\*\*  $p < 10^{-5}$ . Box plots as in **d**. Source data are provided as a Source Data file.

calcium-independent GCaMP response to correct for the hemodynamic component of the calcium-dependent GCaMP responses<sup>69</sup>. We recorded from the right posterior cortex, which gave us access to  $-5$ – $6$  HVAs (Fig. 2a). All the reliably segmented HVAs retinotopically represented the stimulus position in visual space (Supplementary Fig. 1d).

The peak-response maps to the textures and scrambles showed activations almost exclusively in V1 and LM (Fig. 2b). When averaging within the ROIs retinotopically matching the visual stimuli (example blue contours in Fig. 2b), the responses were larger for textures than scrambles both in V1 and LM (Fig. 2c, d) and accordingly, the difference in the peak-response maps resulted in a differential modulation localized primarily in V1 and LM (Fig. 2e). To establish statistical significance, we tested the modulation of each pixel against a null distribution derived from the pre-stimulus period (Fig. 2f), and to determine the significance of an entire visual area, we computed the proportion of significantly modulated pixels in each area within retinotopic ROIs, demonstrating that areas V1 and LM were those most prominently modulated by textures relative to scrambles (Fig. 2g). To compare response modulations between V1 and LM, we computed a texture discriminability measure ( $d'$ ) in retinotopically matched ROIs and found that  $d'$  values in LM were significantly higher than those in V1 (Fig. 2h. V1:  $0.41 \pm 0.05$ , s.e.; LM:  $0.79 \pm 0.05$ ; difference,  $p = 3 \times 10^{-6}$ , paired t test,  $n = 11$  mice).

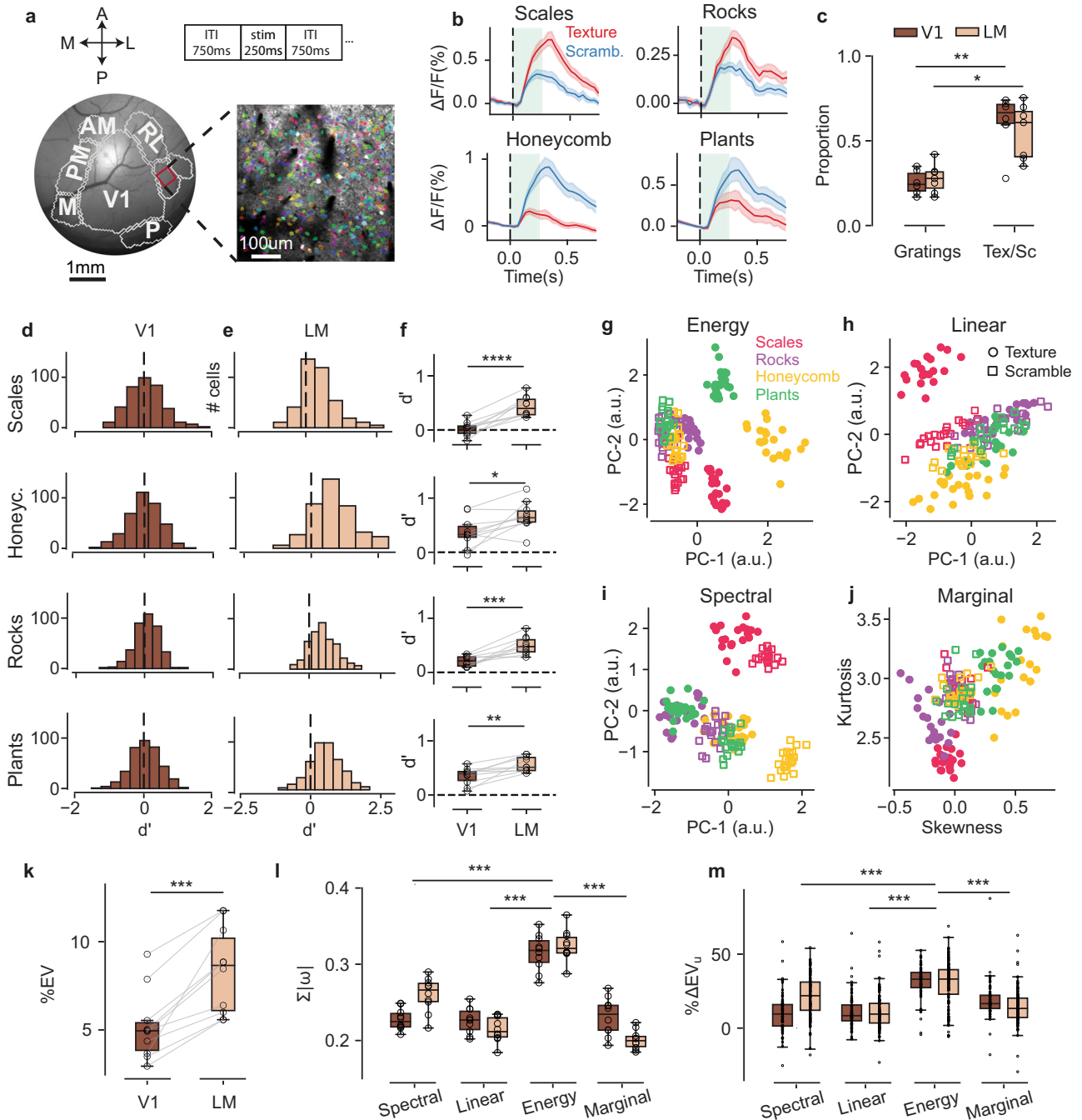
Finally, we observed that, despite the stimuli being retinotopically mapped onto the central-lateral portion of V1 (Fig. 2b, blue contours), significant texture–scramble modulations were most prominent in the

posterior-lateral region of V1 (example in Fig. 2f). To test for a possible representational gradient or asymmetry in spatial representations<sup>70–72</sup>, we performed experiments with full-field texture and scramble images presented monocularly on a monitor sufficiently large as to activate the entire V1 (azimuth,  $[-62.4^\circ, +62.4^\circ]$ ; elevation  $[-48.5^\circ, +48.5^\circ]$ ). Based on maps of elevation and azimuth, we then divided V1 into four quadrants representing the left-right upper and lower visual fields, finding that texture discriminability ( $d'$ ) was consistently higher in the upper visual field (Supplementary Fig. 4).

Together, these results indicate that, at the mesoscopic level, texture selectivity in V1 is biased toward the upper visual field, and when considering a constellation of HVAs surrounding the primary visual cortex, LM is the area with the most significant selectivity to higher-order texture statistics.

### Single-cell responses to texture and scrambles

**Proportion of cells responding to textures in V1 and LM.** We examined the circuit-level representations underlying this mesoscale selectivity using two-photon GCaMP recordings in areas V1 and LM (Fig. 3a). Imaging ROIs (approximately  $530 \mu\text{m} \times 530 \mu\text{m}$ ) in V1 and LM were selected based on the retinotopic coordinates of the visual stimuli, and neural activity was recorded while presenting three classes of visual stimuli: static gratings of different orientations and spatial frequencies (sf; four orientations spaced every 45 degrees,  $100^\circ$  in size, full contrast, sf = [0.02, 0.04, 0.1, 0.2, 0.5] cpd), scrambles and texture images matching the properties of the stimuli used in behavioral



experiments (four families for scrambles and textures, each with 20 exemplars rotated either by 0 or 90 degrees, and with eight repetitions of each image).

The single-cell responses to oriented gratings agreed with what is typically reported in the literature (e.g. refs. 73,74), with approximately 25–30% of the segmented cells being visually responsive (Fig. 3c). The responses to textures and scrambles were rather heterogenous, with some cells strongly responding to textures, others to scrambles, and several showing mixed selectivity (Fig. 3b). In both V1 and LM, there was a significantly larger proportion of cells responding to textures or scrambles relative to gratings (Fig. 3c, V1, textures or scrambles:  $61\% \pm 6\%$ ; LM:  $55\% \pm 6\%$ , s.e.). Despite the significant heterogeneity, responses averaged across cells were significantly larger in LM than in V1 for all texture families (Supplementary Fig. 5a, b). We then quantified the

texture–scramble response modulation of the individual cells using a discriminability measure ( $d'$ ), similar to what was done in mesoscale analyses (Fig. 3d, e), and found that (i) the proportion of cells with significantly positive  $d'$  values (i.e., with larger values in response to textures) were higher in LM than in V1 for all families (Supplementary Fig. 6c); (ii) the average  $d'$  value was higher in LM than V1 for all families (Fig. 3f, V1: average  $d' = 0.24 \pm 0.01$ , LM: average  $d' = 0.54 \pm 0.01$ ,  $p = 2 \times 10^{-4}$ , paired t-test,  $n = 10$ ; Supplementary Fig. 6a,b), which reflected larger response amplitudes to textures than scrambles (Supplementary Fig. 6d).

Together, these results indicate that underlying the increased widefield texture selectivity in LM is both an increase in the proportion of texture-selective cells as well a larger texture–scramble modulation of individual cells.

**Fig. 3 | Single-cell responses in LM better discriminate textures from scrambles.** **a** Multi-area imaging, as in Fig. 2a, with inset showing a representative ROI for two-photon recordings; colored dots indicate the segmented cells responsive to textures and/or scrambles (“stim”, top). **b** Top panels: two example cells responding more strongly to a texture family (red) than scrambles (blue); bottom panels, two example cells for the opposite selectivity. Shaded bands are 95% CIs; center is the mean across 80 texture samples (20 samples per family); vertical broken line indicates the time of stimulus onset, and the green rectangles show the stimulus duration (250 ms). **c** The proportion of cells that significantly responded to oriented gratings in V1 and LM (V1:  $25\% \pm 3\%$  s.e.,  $n = 6$  mice; LM:  $28\% \pm 3\%$ ,  $n = 7$ ; average no. of segmented cells =  $381 \pm 44$  in V1 and  $344 \pm 46$  in LM) was lower than the proportion of cells responding to either textures or scrambles (Tex/Sc; V1:  $61\% \pm 6\%$  s.e., LM:  $55\% \pm 6\%$ ). Gratings vs. Tex/Sc in V1:  $p = 0.002$ , LM:  $p = 0.015$ , two-sided paired *t*-test. Box plots indicate the median with a horizontal bar; the box height denotes the inter-quartile range (IQR, 1st and 3rd quartile) and the whiskers extend by  $1.5 \times$  IQR. Distributions of the texture–scramble discriminability values ( $d'$ ) computed for each cell. Rows are for texture families; **d** for V1, **e** for LM: colors as in **c**. Data from a representative experiment. **f** The mean  $d'$  values for all the experiments in V1 and LM ( $n = 10$  mice, open dots); connecting lines for the same-mouse data; V1  $d'$ : scales =  $0.02 \pm 0.04$  s.e., rocks =  $0.21 \pm 0.03$ , honeycomb =  $0.37 \pm 0.08$ , plants =  $0.35 \pm 0.05$ , LM  $d'$ : scales =  $0.43 \pm 0.05$ , rocks =  $0.50 \pm 0.05$ , honeycomb =  $0.67 \pm 0.08$ , plants =  $0.56 \pm 0.04$ ;  $p$ -values from two-sided paired *t*-tests with Holm-Bonferroni correction: scales,  $p = 7 \times 10^{-5}$ ; rocks,  $p = 6 \times 10^{-4}$ ; honeycomb,  $p = 2 \times 10^{-2}$ ; plants,  $p = 4 \times 10^{-3}$  ( $n = 10$ ). Box plots as in **c**. **g–j** Two-dimensional PCA embedding of each of the four groups of image statistics (titles). The dots indicate texture exemplars (20), and the squares scramble exemplars (20). Color code for

texture families in the legend. The same images were used for both behavioral and imaging experiments. **k** The explained variance (EV, %) by the encoding linear model based on PS image statistics, comparing V1 to LM; only cells for which  $EV \geq 1\%$  have been included in the analysis (permutation test, Methods); each dot is a mouse; connecting lines for the same-mouse data; V1:  $5.3\% \pm 0.6\%$  s.e., LM:  $8.2\% \pm 0.7\%$ , cross-validated,  $n = 10$ , V1 – LM difference,  $p = 7 \times 10^{-4}$ , two-sided paired *t*-test. Box plots as in **c**. **l** The sum of weight values for each of the PS statistic groups of the fitted regressive model; each dot is an average across cells for a given mouse. The energy statistics are significantly higher than all others, both in V1 and in LM:  $p < 0.001$  for energy compared to all other statistics; V1 comparisons: energy vs. linear:  $5 \times 10^{-8}$ ; energy vs. marginal:  $1 \times 10^{-6}$ ; energy vs. spectral:  $1 \times 10^{-6}$ ; linear vs. marginal: 0.7; linear vs. spectral: 0.7; marginal vs. spectral: 0.9. LM comparisons: energy vs. linear:  $1 \times 10^{-12}$ ; energy vs. marginal:  $2 \times 10^{-13}$ ; energy vs. spectral:  $3 \times 10^{-6}$ ; linear vs. marginal: 0.8; linear vs. spectral:  $5 \times 10^{-5}$ ; marginal vs. spectral:  $4 \times 10^{-6}$ .: one-way ANOVA with post-hoc analysis (Tukey HSD). Colors and box plots as in **c**. **m** The “unique” EV ( $\Delta EV_u$ ) for all four PS statistics groups. The cells with a high explained variance by the full model ( $EV \geq 10\%$ ) were included in the analysis. Each dot is the change in explained variance for a cell when using the “full” model or a model missing a given PS statistic. The energy statistics are significantly higher than all others, both in V1 and LM:  $p < 0.001$  for energy larger than all other statistics; V1 comparisons: energy vs. linear:  $< 10^{-12}$ ; energy vs. marginal:  $7 \times 10^{-12}$ ; energy vs. spectral:  $< 10^{-12}$ ; linear vs. marginal:  $1 \times 10^{-4}$ ; linear vs. spectral: 0.9; marginal vs. spectral:  $1 \times 10^{-4}$ . LM comparisons: energy vs. linear:  $5 \times 10^{-15}$ ; energy vs. marginal:  $5 \times 10^{-15}$ ; energy vs. spectral:  $8 \times 10^{-8}$ ; linear vs. marginal: 0.04; linear vs. spectral:  $3 \times 10^{-11}$ ; marginal vs. spectral:  $6 \times 10^{-5}$ ; one-way ANOVA with post-hoc analysis (Tukey HSD). Box plots as in **c**. Source data are provided as a Source Data file.

**Encoding linear model of neural responses.** To isolate the set of statistical features that most prominently drove the texture–scramble selectivity in V1 and LM, we used a previously described mathematical model to parametrize image statistics: the Portilla–Simoncelli statistical model (henceforth, PS model and statistics<sup>15</sup>). This model employs a set of analytical equations to compute the correlations across a set of filters tuned to different image scales and orientations. These statistics can be divided into four main groups: marginal (skewness and kurtosis of the pixel histogram), spectral, linear cross-correlation, and energy cross-correlation statistics. In its complete formulation, the model provides a very high dimensional parametrization of the stimuli (740 parameters), resulting in more parameters than the total number of images (320). Therefore, dimensionality reduction of each PS group of statistics can provide a reduced representation without significant loss in parametrization power (Supplementary Fig. 8, 9). We used Principal Component Analysis (PCA), finding that with even the first two principal components, the energy statistics are best at separating textures from scrambles and between texture types (Fig. 3g–j, Supplementary Fig. 7, 9), as also reported in human psychophysical studies<sup>34</sup>.

Using PS statistics as features, we created an encoding linear model for single-cell responses in V1 and LM. The model’s task was to predict the response of a particular neuron to all texture and scramble exemplars as a weighted linear sum of PS coefficients. When considering the cells for which the model could explain at least 1% of the response variance—that is, a threshold value for the significance of the model’s fits derived from a permutation test (Methods)—we found that the proportion of these cells was higher in LM than in V1 (V1:  $58\% \pm 8\%$  s.e., LM:  $78\% \pm 3\%$ ,  $p = 1.0 \times 10^{-4}$ , paired *t*-test,  $n = 10$ ), with a higher average explained variance in LM (Fig. 3k; Supplementary Fig. 6e). The energy cross-correlation statistics had the largest contribution to the explained variance (Fig. 3l), which was also confirmed by an analysis of “unique” variance explained<sup>75</sup> (withholding a particular group of PS statistics), and it was found that the energy cross-correlation statistics was again the main contributor (Fig. 3m).

As these results show, underlying the increased selectivity for textures in area LM and a larger proportion of cells having such selectivity is a stronger responsiveness to statistical features that are texture-defining—that is, those quantified by the energy cross-correlation PS statistics.

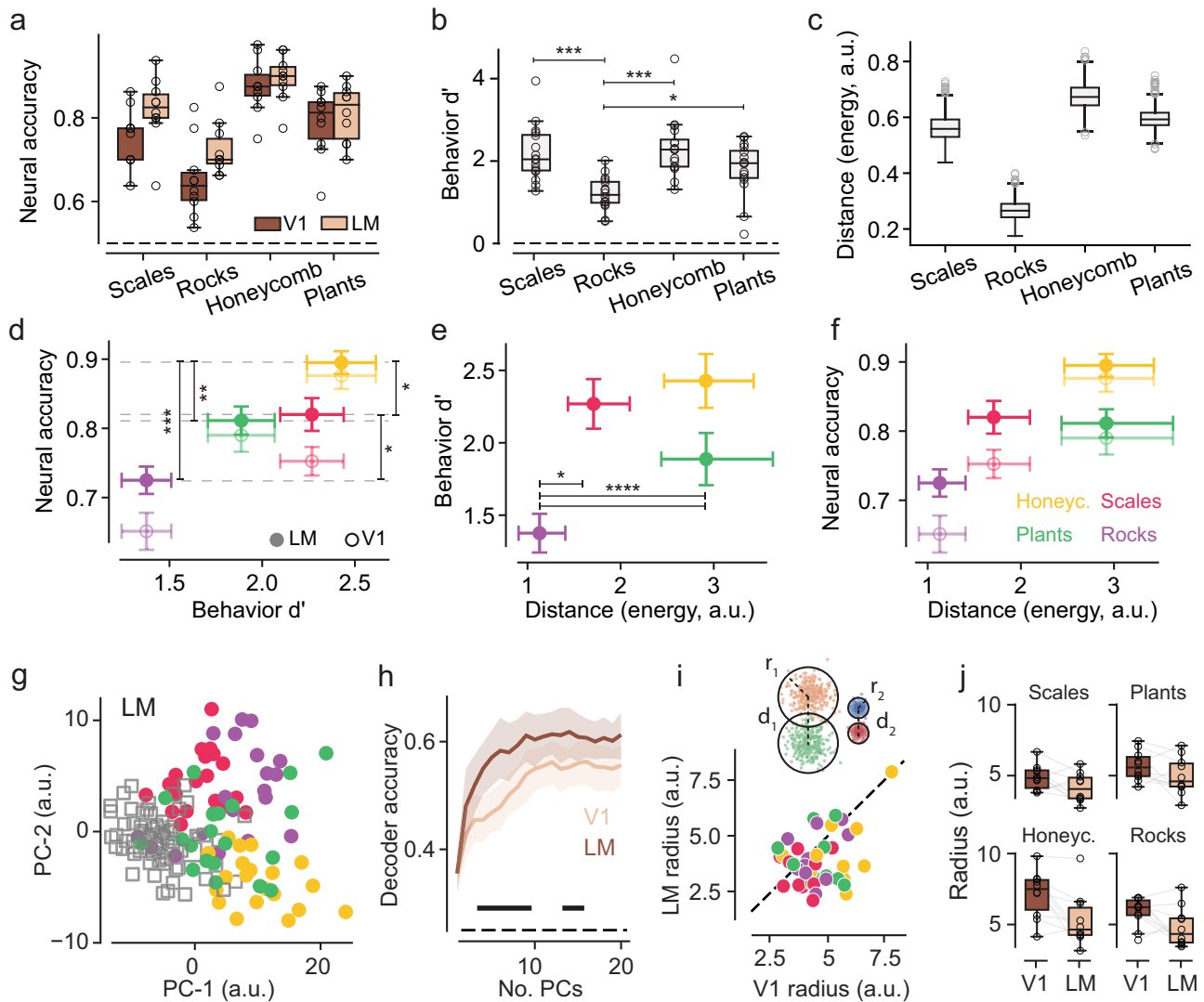
### Population responses to texture images

Next, we examined whether we could identify signatures of texture selectivity, more significantly so in LM than in V1 at the level of population encoding. To discriminate the activity of the texture–scramble pairs in V1 and LM, we trained a binary logistic classifier. The decoder was largely above the chance level (50%) for all pairs (Fig. 4a), with significantly larger performance in LM than in V1 when grouping all the texture families (V1,  $77\% \pm 1\%$  (s.e.); LM,  $81\% \pm 2\%$ ,  $p = 0.007$ , paired *t*-test,  $n = 10$  mice). In both V1 and LM, the rocks family was the one with the lowest classification accuracy (Fig. 4a,  $p = 4 \times 10^{-7}$ , one-way ANOVA; performance of rocks different from all pairs, repeated measures correction,  $p < 0.035$ , post-hoc Tukey HSD test,  $n = 10$ ). Notably, a similar drop in performance was also observed in the  $d'$  measures of behavioral performance, where the lowest performance was observed for this texture–scramble pair consistently in individual mice trained across all four texture–scramble pairs, and across animals (Fig. 4b,  $p = 3 \times 10^{-5}$ , one-way ANOVA; repeated measures correction,  $p < 0.03$ , post-hoc Tukey HSD test,  $n = 16$ ).

### Linking image statistics to neural and behavioral representations.

Next, we examined whether the correlation in neural and behavioral discriminability could be related to the statistics of the images. For instance, if the statistics of the rock exemplars were particularly similar to those of their scrambles compared to other families, then this reduced statistical discriminability may explain the drop in both behavioral and neural discriminability. We thus defined a distance metric in a statistical stimulus space based on a reduced set of PS statistics (Fig. 3g). In each subspace, we measured the inter-cluster distances (normalized by the clusters’ spread) between the textures and the corresponding scrambles, finding that the rocks family had a significantly smaller texture–scramble distance than the other families in the energy statistics subspace (Fig. 4c, e). For the other statistical subspaces, although the texture–scramble distances of rocks were still the shortest compared to the other families, they overlapped with at least one other family (Supplementary Fig. 10).

Together, the correlation between the PS-distance metric in the energy subspace, which best captures texture-defining statistics, and the drop in neural decoding and behavioral performance associated with the rocks family, suggest a tight linking framework between high-



**Fig. 4 | Statistical, behavioral, and neural discriminability correlate with the geometry of texture representations in LM.**

**a** Accuracy of a linear classifier trained to discriminate textures from the scrambles for all four families using the neural responses from LM-ROI and V1-ROI. Each dot indicates a mouse. The decoder accuracy is above 50% chance level for all pairs (V1: scales  $75\% \pm 2\%$  s.e.,  $p = 1 \times 10^{-6}$ , V1: rocks  $65\% \pm 3\%$ ,  $p = 3.5 \times 10^{-4}$ , V1: honeycomb  $88\% \pm 2\%$ ,  $p = 2 \times 10^{-8}$ , V1: plants  $79\% \pm 2\%$ ,  $p = 1.1 \times 10^{-6}$ , LM: scales  $82\% \pm 2\%$ ,  $p = 2 \times 10^{-7}$ , LM: rocks  $72\% \pm 2\%$ ,  $p = 1.9 \times 10^{-6}$ , LM: honeycomb  $89\% \pm 2\%$ ,  $p = 3.0 \times 10^{-9}$ , LM: plants  $81\% \pm 2\%$ ,  $p = 1.5 \times 10^{-7}$ ; one-sample *t*-test,  $n = 10$  mice; ANOVA values reported in **d**) Box plots indicate the median with a horizontal bar; the box height denotes the inter-quartile range (IQR, 1st and 3rd quartile) and the whiskers extend by  $1.5 \times$  IQR. **b** Behavioral discriminability ( $d'$ ), as in Fig. 1f, but for a subset of the mice that completed the texture-scramble tasks for all four families ( $n = 16$ , open dots;  $p = 2.0 \times 10^{-5}$ , one-way ANOVA across families; post-hoc analysis (Tukey HSD): rocks vs. honeycomb:  $p < 0.0001$ , rocks vs. plants:  $p = 0.04$ , rocks vs. scales:  $p = 0.0002$ ). Box plots as in **a**. **c** Normalized distances for each texture/scramble family pair for the energy cross-correlation statistics of the images; gray dots indicate outliers from a bootstrapping procedure (Methods). Box plots as in **a**. **d** The combined 2D plot from the data in **a** and **b**; the error bars are s.e.; colors as in **f**; reference gray horizontal broken lines for pairwise statistical comparisons for area LM (filled dots), one-way ANOVA for texture families:  $p = 3.7 \times 10^{-5}$ ; post-hoc analysis (Tukey-HSD): honeycomb-plants:  $p = 0.04$ ; honeycomb-rocks:  $p = 1.2 \times 10^{-5}$ ; plants-rocks:  $p = 0.03$ ; rocks-scales:  $p = 0.016$ . V1 classifier, one-way ANOVA for texture families (asterisks not shown):  $p = 1.7 \times 10^{-6}$ ; post-hoc analysis: honeycomb-rocks:  $p = 6 \times 10^{-7}$ ; honeycomb-scales:  $p = 0.005$ ; plants-rocks:  $p = 0.0013$ ; rocks-scales:  $p = 0.009$ ;  $n = 10$ , mean accuracy for each mouse. **e** Behavioral discriminability as in **b** plotted against the inter-cluster distances (**c**) for each texture/scramble family pair and for the energy cross-

correlation statistics. The error bars are s.e. for the behavioral data and bootstrap confidence intervals with Šidák correction for multiple comparisons ( $\alpha = 0.05$ ) for statistical distance: CIs = 99.15,  $p_{\text{boot}} < 0.05$ , for all pairwise comparisons with rocks;  $n = 1000$  bootstraps. **f** Neural classifier accuracy, as in **a**, against the inter-cluster distances for the energy cross-correlation statistics (**c**);  $n = 10$  mice. **g** 2D scatter plot for the first two PCA components of the neural responses from LM-ROI for one example animal; each dot is an exemplar (averaged across repeats, image rotations, and time frames around the peak response); filled circles for textures and empty squares for scrambles; colors as in **f**. **h** Accuracy of a multinomial classifier ( $n = 10$  mice) discriminating between the texture families as a function of the number of components, separately in V1 and LM PCA spaces. The shaded regions correspond to the 95% confidence intervals across all mice; the center is the mean accuracy of the classifier (i.e., the mean of the 5-fold cross validation procedure) across all 10 mice. The black horizontal bars indicate the range of PCA components for which the classifier accuracy is statistically different between V1 and LM (paired *t*-test,  $p$ -values  $< 0.05$ ). **i** Top: schematic plot illustrating the metrics used in the neural PCA space. For every cloud of points in the PCA space, we measure its radius (e.g.,  $r_1$ ,  $r_2$ ) and its distance with respect to another cloud (e.g.,  $d_1$ ,  $d_2$ ). The clouds on the left show larger radii and inter-cluster distance compared to the clouds on the right. Bottom: scatter plot of the cluster radii in V1 (x-axis) and LM (y-axis) for all mice ( $n = 10$ ). Each dot is a cluster radius for a given texture family and mouse; colors as in **f**. The black dotted line is the diagonal. **j** Radius values in V1 and LM; each dot corresponds to a particular mouse and each panel a different texture family ( $n = 40$ , 10 mice  $\times$  4 families), gray lines are paired animals and families between V1 and LM;  $p = 0.04$ , repeated-measures ANOVA brain-area effect;  $p = 6.0 \times 10^{-6}$  family effect,  $p = 0.002$  interaction effect. Box plots as in **a**. Source data are provided as a Source Data file.

order image statistics, population encoding in V1 and LM, and behavioral performance (Fig. 4d–f).

**V1 and LM differences in the representational geometry of texture families.** The results from the binary logistic classifier trained to discriminate between the texture–scramble pairs suggest representational differences between V1 and LM. For instance, significantly fewer principal components (PCs) were needed in LM to attain maximum performance (two to four dimensions) whereas V1 required thrice as many, between four and 12 PCs (Supplementary Fig. 11). Further evidence for representational differences between V1 and LM was provided by a decoding analysis attempting to discriminate between texture families from the neural activity in V1 and LM. We used a multinomial logistic classifier trained to categorize the four texture families across each of the 40 exemplar images for each family. Since the number of cells differed across experiments, we used PCA to fix the representational dimensionality of the activity space. Even with only two PCA components, the collective activations of the visually responsive cells across all texture and scramble stimuli already formed separate activity subspaces (or “clusters”, Fig. 4g) with an average explained variance above 15% (V1:  $15.5\% \pm 1.4\%$  s.e. LM:  $19.1\% \pm 1.2\%$ ). The cross-validated classifier performed significantly above chance level in both areas, plateauing at approximately 60% performance with  $\sim 10$  PCA components (Fig. 4h). The LM decoder outperformed the V1 decoder, with significant differences observed reliably in the range between two and sixteen PCA components (Fig. 4h).

To highlight the properties of the population encoding that could explain the increased classification performance in LM, we studied the geometry of texture representations in a shared 16-dimensional PCA space of V1 and LM activations, in which the texture–texture decoder had the largest (significant) discriminability power (Fig. 4h). Each point in this space corresponded to a texture exemplar (averaged across repeats) labeled according to the corresponding texture family (2D schematic of the 16D representations in Fig. 4i). For every family pair (40 exemplars per family) we computed a Mahalanobis distance measure, which demonstrated an overall increase in distance in LM compared to V1, in agreement with the performance of the multinomial classifier ( $24 \pm 5\%$  distance increase in LM vs. V1,  $p = 0.002$  paired t-test,  $n = 60$ , 10 mice  $\times$  6 pairs). Both the classifier and the Mahalanobis distance measure are sensitive to the relative “shapes” of the underlying distributions. To test for simple geometrical changes from V1 to LM, for every family we computed the spread of the activations associated with the 40 exemplars—that is, the radii of the activity subspaces and their pairwise distances (“inter-cluster” Euclidean distances). In LM, we found that the cluster radii were significantly smaller than in V1 (Fig. 4i,j; repeated-measures ANOVA: interaction effect,  $p = 0.002$ ; brain-area effect,  $p = 0.04$ ; family effect,  $p = 6.0 \times 10^{-6}$ ), with no evidence for smaller inter-cluster distances in LM compared to V1.

In conclusion, a population-level signature of the increased selectivity for energy cross-correlation statistics in LM is a change in the representational geometry of the texture stimuli with LM having more “compact” representations than V1, as evidenced by the smaller subspace radii. These findings suggest that the more compact representations in LM contribute to the increased classification performance observed in this area.

## Discussion

We found that mice can perceptually detect higher-order statistical dependencies in texture images, discriminating between textures and scrambles and between different texture families. Across visual areas, V1 and LM were those most prominently selective to texture statistics, with LM more so than V1, significantly driven by the energy cross-correlation image statistics. The representational geometry of

population responses demonstrated subspaces for each texture–scramble pair, with better stimulus decoding in LM than in V1. The distances between the texture–scramble subspaces changed according to the stimulus statistical dependencies, more significantly in the energy cross-correlation statistical components. The textures statistically most similar to scrambles (i.e., exemplars from the rocks family) had the shortest distances between the corresponding neural subspaces, with the worst perceptual discriminability by the animals and by a decoder trained on the neural representations. This was observed consistently in the animals trained on various texture–scramble pairs as well as across animals for this specific pair. Finally, the neural representations for different texture families were also easier to discriminate in LM than in V1, with LM having more compact subspaces (smaller radii) for individual textures.

Efficiency, in reference to the efficient coding hypothesis<sup>22</sup>, highlights a correspondence between input statistics, perceptual sensitivity, and the allocation of computational (and metabolic) resources. A neural code is efficient if it can reflect environmental statistics; such a code will favor basic visual features that are more common, relying on non-uniform neural representations and perceptual sensitivity<sup>23,24,26,27</sup>. This implies a close correspondence between neural, perceptual, and statistical representations. We studied this correspondence by examining the geometry of such representations in V1 and LM and identifying “rocks” as the family most similar to its scramble exemplars, with neural-distance representations and behavioral performance also being the smallest for this family. This was reliably observed in animals (tested across various texture–scramble pairs) and across animals for this pair. The selected texture families were chosen because of their likely ethological relevance to mice (e.g., rocks and plants) and their extensive use and characterization in the texture literature<sup>38,59</sup>. They also had sufficiently diverse statistical dependencies to permit a simple statistical similarity ranking between the texture–scramble pairs. However, future work could adopt a more principled approach in selecting texture families based on the statistical distance measure, as adopted in this study. This would allow us to define a psychometric difficulty axis in the stimulus–statistics space to be explored parametrically, both for texture–scramble and texture–texture discrimination. For the latter in particular, this approach could overcome a current statistical limitation: the six texture–family pairs span a relatively narrow range of distances in stimulus statistics, requiring an extremely large number of trials to test for differences in behavioral performance and neural representations, both within and across mice. Texture synthesis guided by a predetermined sampling of the relevant distances along a psychometric difficulty axis could ease the burden of collecting an exceedingly large dataset.

To examine the perceptual ability of mice to discriminate textures, we carefully controlled the stimulus statistics of each exemplar. We customized a CNN-based approach for texture synthesis to achieve the equalization of lower (e.g., luminance, contrast, and marginal PS statistics) and higher statistical dependencies (e.g., linear and energy cross-features PS statistics). Further, we normalized the power spectrum in a frequency band of high perceptual sensitivity for mice and generated several metameric exemplars<sup>43</sup> differing in pixel-level representations but otherwise having identical statistical dependencies. We also introduced image rotations to ensure that the animals could generalize along this stimulus dimension. Finally, we tested the trained animals with new sets of metameric exemplars, confirming that “brute force” memorization of low-level features was not used in the task<sup>67</sup>. This approach gave us control over which statistical features the mice could use in the task and which component is critical when linking the statistical dependencies of the visual stimuli to neural and perceptual representations. In this respect, our approach may be preferable to using *synthetic* textures, in which typically only a reduced set of statistics of interest is under parametric control, while others are left free to (co)vary<sup>13,14,76–78</sup>.



The linking framework between stimulus statistics, neural representations, and perceptual sensitivity was most significant for the energy cross-correlation statistics. These statistics capture dependencies in high-order spatial correlations, sampling different parts of the image with filters having different spatial scales and orientations<sup>15</sup>. Therefore, energy statistics are sensitive to the “ergodic” properties of textures, that is the homogeneity of the statistics across different parts of an image or, for a given part, across exemplars from the same texture family<sup>1</sup>. Spatial structures and patterns repeated across the image (e.g., elongated contour bands) are absent in scramble images, which however are matched in average orientation power to texture images. It is possible that mice rely on these patterns in the behavioral tasks. Previous psychophysical studies in humans<sup>34</sup> have shown that energy components play a crucial role in predicting perceptual sensitivity to texture images, motivating the use of perceptual components in models of texture synthesis<sup>39</sup>. However, bridging the gap between the perceptual strategies employed by mice in our tasks and the link between energy statistics and human perception remains challenging. Finally, additional research is necessary to extend our findings to non-ergodic natural images, such as images of objects and faces, also characterized by high-order spatial correlation properties.

The asymmetry in texture-scramble discriminability ( $d'$ ) observed in V1, with a bias for the upper visual field, could reflect an adaptive mechanism to natural-image statistics. Previous studies have identified gradients in mouse V1 related to variables such as binocular disparity<sup>70</sup>, coherent motion<sup>71</sup>, and UV-green color contrast<sup>72</sup>. These gradients have been related to statistical properties in visually relevant environments. Likewise, the observed gradient in the discriminability measure ( $d'$ ) may signify a heightened sensitivity to high-order spatial correlations in the upper visual field associated with natural elements in the visual scene—not uniquely textural—such as predators or landmarks used for navigation.

The prominent texture selectivity found in area LM is consistent with what is known about the area specialization of the mouse visual cortex, implicating LM in the processing of content-related (semantic) visual information<sup>74,79–86</sup>, in high-fidelity representations of spatial features—including those of textures<sup>58</sup> and with inactivation studies demonstrating the necessity of LM for the perception of even simple visual stimuli<sup>80,83</sup>.

At the circuit level, an analysis of the representational geometry of LM population responses<sup>87,88</sup> revealed distinct activity subspaces associated with different texture families. These texture “manifolds” are reminiscent of the concept of object manifolds introduced in relation to the processing of complex objects along the ventral stream in primates<sup>89–92</sup> and in mice<sup>57,93</sup>. When comparing LM to V1 representations, we found a reduction in the size (radius) of texture clusters, with this effect leading to an overall improved linear discriminability of texture families in LM compared to V1. One interpretation is that the increased discriminability from V1 to LM is related to an increase in the representational invariances to image statistics, as suggested by previous studies on rats<sup>55,56</sup> and mice<sup>57</sup>. The reduction in cluster sizes reflects an overall more compact representation of the four texture families, which may relate to LM achieving a higher encoding capacity than V1 while, at the same time, retaining large encoding accuracy for textures. Another possibility is that the V1 texture representations reflect an “incomplete inheritance” from LM via top-down signal processing<sup>94</sup>; experiments inactivating LM while recording from V1 could elucidate this point.

Neural recordings were done in untrained animals passively viewing the stimuli, thus enabling comparisons with primate studies that used similar preparations<sup>34,36,95</sup>. Furthermore, neural recordings in untrained animals eliminate the possibility that the observed selectivity and representational features emerge as a consequence of the task-learning process. Rather, our analyses likely highlight a computational property of the visual system emerging from an evolutionarily

refined genetic program<sup>28</sup> and from exposure to a rich set of image statistics during development. The observation that in naïve animals the decoding quality of the neural signals follows the statistical separability of texture–scramble images, mirrored by congruent performance modulations in trained animals, supports this interpretation. It is also conceivable that learning and attentional processes, as animals engage in tasks, might affect the properties of neural representations<sup>1,96,97</sup>. Therefore, in future studies, it would be of interest to examine the neural dynamics underlying texture representations during the different phases of learning.

In conclusion, our results demonstrate the signal processing of naturalistic stimuli in the mouse visual cortex akin to what has been observed in primates, additionally highlighting an intimate link between the geometry of neural representations, stimulus statistical dependencies, and perceptual behavior, which is a distinct hallmark of efficient coding principles of information processing. Considering that similar processing features are also found in V2/LM equivalents in artificial neural networks, our results likely reflect a general efficient coding principle emerging in hierarchically organized computational architectures devoted to the extraction of semantic information from the visual scene.

## Methods

### Subjects

All procedures were reviewed and approved by the Animal Care and Use Committees of the RIKEN Center for Brain Science. The behavioral data for the texture–scramble and texture–texture discrimination visual task were collected from a total of 21 mice: six Camk1TA;TREGCaMP6s (four males and two females), 14 C57BL/6J WT (11 males, three females), and one male CaMKII $\alpha$ -Cre. For the passive imaging experiments, we used a total of 11 mice (11 for widefield and 10 for two-photon): six CaMKII $\alpha$ -Cre transgenic mice (four males and two females) and five C57BL/6J WT (two males and three females). The age of the animals typically ranged between eight and 28 weeks old from the beginning to the end of the experiments. The mice were housed under a 12–12 h light–dark cycle. Temperature was kept in the 20–24 °C range and humidity at 45–60%.

### Cranial window implantation

As described in ref. 64, for the implantation of a head-post and optical chamber, the animals were anesthetized with gas anesthesia (Isoflurane 1.5–2.5%; Pfizer) and injected with an antibiotic (Baytril, 0.5 ml, 2%; Bayer Yakuin), a steroidal anti-inflammatory drug (Dexamethasone; Kyoritsu Seiyaku), an anti-edema agent (Glyceol, 100  $\mu$ l; Chugai Pharmaceutical) to reduce brain swelling, and a painkiller (Lepetan, Otsuka Pharmaceutical). The scalp and periosteum were retracted, exposing the skull, and then a 5.5 mm diameter trephination was made with a micro drill (Meisinger LLC). Two 5 mm coverslips (120–170  $\mu$ m thickness) were positioned in the center of the craniotomy in direct contact with the meninges, topped by a 6 mm diameter coverslip with the same thickness. When needed, Gelform (Pfizer) was applied around the 5 mm coverslip to stop any bleeding. The 6-mm coverslip was fixed to the bone with cyanoacrylic glue (Aron Alpha, Toagosei). A round metal chamber (7.1 mm diameter) combined with a head-post was centered on the craniotomy and cemented to the bone with dental adhesive (Super-Bond C&B, Sun Medical), which was mixed with a black dye for improved light absorbance during imaging.

### Viral injections

For imaging experiments, we injected the viral vector rAAV1-syn-jGCaMP7f-WPRE ( $4 \times 10^{12}$  gc/ml, 1000 nl) into the mice’s right visual cortex (AP,  $-3.3$  mm; LM 2.4 mm from the bregma) at a flow rate of 50 nl/min using a Nanoject II (Drummond Scientific, Broomall, Pennsylvania, USA). The injection depth was 400  $\mu$ m. After confirmation of

fluorescent protein expression (approximately two weeks after the AAV injection), we made a craniotomy (5.5 mm diameter) centered on the injection site while keeping the dura membrane intact and implanted a cover-glass window, as described above.

## Behavior

**Behavioral training procedure.** Water-restricted mice were habituated to our automated behavioral training setups with self-head fixation, as previously described<sup>64</sup>. The training of mice progressed according to four stages with increasing difficulty, both procedural and perceptual, and with the fourth stage involving the final tasks described in the Results section (both the texture–scramble and texture–texture tasks). In the first stage, trial timing and stimulus properties were already set as in the final stage (Fig. 1d). However, 1) the “go” stimuli were shown in 70% of the trials (instead of 50% as in the fourth stage); 2) the minimum wheel rotation required to trigger a response was 5° instead of 45°; 3) the maximum wheel rotation that was allowed during the last second of the ITI was larger (20° instead of 5°); 4) the reward size was 8  $\mu$ l instead of 4  $\mu$ l. During this training stage, mice learned the association between wheel rotation and water reward contingent on the stimulus presentation on the screen. After they learned to rotate the wheel contingent to stimulus presentation in at least 80% of the trials for three consecutive sessions, they were moved to the second training stage, with the following changes: (1) the “go” stimuli were shown in 70% of the trials; (2) the wheel rotation angle to signal a response was increased to 15°; (3) the maximum wheel rotation allowed during the ITI was decreased to 5°; and (4) the water reward was lowered to 4  $\mu$ l. After the mice reached at least 70% hits for three consecutive sessions, they were moved to the third training stage, in which the only change was an increase in the wheel rotation angle to 30° to signal a response. After reaching at least 70% hits for three consecutive sessions, the mice were moved to the fourth and final training stage with 50% hit trials. Most of the mice started the training with the honeycomb or scales texture/scramble family. Afterwards, we randomly selected the next family until all four families were successfully discriminated against the corresponding scrambles. A texture–scramble family discrimination was considered completed when the mouse had a  $d' > 1$  consistently over 10 consecutive sessions. The training details for the texture–texture task are described in the “Texture–texture task” section. In the final stage, mice received 4–5 ml of water daily. In preceding stages, in cases where mice failed to acquire sufficient hydration to maintain a healthy body weight (measured daily), specifically falling below 75% of the baseline body weight, we administered a significant bolus of water gel after the session. Alternatively, we temporarily withdrew the animal from the training protocol in instances where the mouse struggled to rapidly regain a healthy weight. Mice were typically not trained on weekends when they had free access to water.

**Behavioral performance.** We evaluated behavioral performance using the discrimination metric  $d'$  from signal detection theory<sup>98</sup> which is defined as:  $d' = Z(\text{hit-rate}) - Z(\text{false-alarm-rate})$ , where  $Z$  is the inverse cumulative normal distribution function, “hit rate” is the proportion of correct “go” trials, while “false alarms” is the proportion of “no go” trials with erroneous responses.

**Texture–scramble task.** Mice independently fixed their headplate to the latching device twice a day in a fully automated behavioral setup<sup>64</sup> that was connected to their home-cage. It comprised a self-latching stage, a rubber wheel with a quadrature encoder sensor to read the wheel's position<sup>99</sup>, a spout that dispensed water drops (4  $\mu$ l), and a computer monitor positioned in front of the latching stage. Mice were required to rotate the toy wheel with their front paws contingent on a texture stimulus shown on the screen (the “hit” trials were rewarded with a water drop; the “false alarm” responses were discouraged by

presenting a full-field flickering checkerboard pattern for 10 seconds; no feedback was given for “misses” and “correct rejects”). Regarding the temporal structure of the trial (shown in Fig. 1d), a session began with an ITI with an isoluminant gray screen (with the same mean luminance level of the texture and scramble images). The ITIs lasted for four to six seconds chosen from a randomly uniform distribution. Mice had to refrain from rotating the wheel, with movements during a one-second period before the onset of the visual stimulus extending the ITI by one second. The stimuli had a 50% chance of being either a go stimulus (texture exemplar) or a no-go stimulus (scramble exemplar). The parameters of the stimuli matched those used in the imaging experiments: 100° in visual angle, with a raised cosine mask to reduce sharp edges (high-frequency components), and the texture family to be discriminated was kept constant during the entire session, randomly selecting the image to be displayed in each trial from a set of 20 exemplars. Following the stimulus presentation, the mice had two seconds to respond (response window). A wheel rotation was counted as a response if it exceeded 45°. After a hit trial, a water reward was given, which was followed by a one-second period, during which the stimulus remained visible on the screen, which then disappeared at the beginning of the ITI period with a randomized four to six second duration. In false-alarm trials, the stimulus disappeared after the wheel rotation, and a flickering checkerboard pattern (2 Hz) was displayed for 10 seconds followed by an ITI period. For miss trials, a new ITI began at the end of the two-second response window. The session ended either when the mice received 400  $\mu$ l of water or when the session's duration reached 1800 seconds. To verify that the mice did not rely on “brute force” memorization of the luminance patterns shown on the screen to solve the task<sup>100</sup>, in a subset of expert animals ( $n = 17$ ), trained on all four texture–scramble family pairs, we introduced new sets of texture and scramble exemplars (20 each) and compared the performance of mice in the five sessions before and after the change in exemplars.

**Texture–texture task.** The mice trained in the texture–texture go/no go task were both a subset of the mice trained in the texture–scramble ( $n = 14$ ) and a new cohort of naïve mice ( $n = 2$ ). If the mouse had been previously trained in the texture–scramble task expert, we simply changed the protocol so that a randomly chosen texture family (20 exemplars) was the new “go” stimuli and, similarly, another randomly chosen texture family (20 exemplars) was the new “no go” stimuli. Instead, for naïve mice, we trained them following the same training procedure described for the texture–scramble task but using exemplars from another randomly chosen texture family instead of scrambles.

## Image synthesis

**Texture synthesis.** As described in ref. 59, convolutional neural networks (CNNs) can be used to extract a compact representation of texture images by measuring the activation patterns of a CNN to a given texture. These activations are an over complete multi-scale representation<sup>59,101</sup> that can be used to synthesize an arbitrary set of texture exemplars. Specifically, the first step for the synthesis of a novel texture exemplar relative to a reference texture (“target”,  $x$ ) is to obtain a CNN parametrization of  $x$ —that is, its feature vector representation,  $f(x)$ . This is done by concatenating the spatial means of the feature-map activations in each of the five VGG16 layers, which results in a feature vector of size 1,472:  $f(x) = \{\mu_{x_j}^{(i)}; i = 1, \dots, m; j = 1, \dots, n_i\}$ , where  $m = 5$  is the number of convolutional layers,  $n_i$  is the number of feature maps in convolutional layer  $i$ ,  $\bar{x}_j^i$  is the spatial mean across filter activations in the feature map  $j$ , and  $\mu_{x_j}^{(i)}$  is the set of such means for layer  $i$ . The second step is to obtain a feature vector representation,  $f(y)$ , of a Gaussian-noise image  $y$ :

$f(y) = \{\mu_y^{(i)}; i=1, \dots, m; j=1, \dots, n_i\}$ . To obtain  $f(x) \approx f(y)$ , we solve an optimization problem (with an L1 loss):

$$y^* = \operatorname{argmin}_y \sum |f(x) - f(y)| \quad (1)$$

Where  $y^*$  is the fully optimized image relative to the target image,  $x$ .

This approach is nearly identical to that of ref. 59, with the only difference being that we did not add the mean of the three-color channels of  $x$  to our feature transform<sup>39</sup> since in our framework it created some degree of “pixelation” in the synthesized images. Rather, as an additional step after optimization, we normalized the images to have equal mean luminance and standard deviation (RMS contrast), as detailed below.

For the synthesis of the textures in Supplementary Fig. 7 we used the Portilla-Simoncelli algorithm to generate images for four texture families. First, we computed the PS statistics of four original textures (using randomly chosen exemplars), and then we generated new textures constraining linear, marginal, and spectral statistics to be those of the “scales” family. Instead, the energy statistics were the original statistics of each of the four texture families (MATLAB function “TextureSynthesis.m”, <http://www.cns.nyu.edu/~lcv/texture/>).

**Texture normalization.** To ensure that texture exemplars had the same lower-order statistics (mean luminance and RMS contrast), we z-scored the pixel intensity values, multiplied them by a fixed contrast (standard deviation,  $\sigma = 0.15$ ), and, finally, added a fixed mean luminance value ( $\mu = 0.5$ ). This normalization was applied to all the “target” texture images (relative to the synthesis procedure with VGG16) and the synthesized texture exemplars, as there were small differences in the luminance and contrast relative to the target after each exemplar was synthesized. Furthermore, to ensure that the spatial frequency content of the textures was within the range of mouse perceptual sensitivity, we used an iterative algorithm in which we progressively rescaled the “target” texture images such that 1) >95% of the spatial-frequency amplitudes of all the target textures lied within the 0.0 to 0.5 cpd<sup>102</sup> interval; 2) the average amplitude spectrum overlapped across families in the frequency range between 0.01 and 0.5 cpd.

**Scramble generation.** Scrambles are the noise images spectrally matched to the textures<sup>34</sup> generated via FFT-transform of a given texture exemplar (changing for different texture–scramble pairs) and randomizing the phase components while keeping the amplitude ones. Phase randomization was done by drawing the phase values from an FFT-transform of a Gaussian-noise image. The thus-generated scrambles retained the same average orientation and spatial-frequency power as the texture exemplars but lacked the higher-order statistical dependence of the textures<sup>34</sup>. For each of the synthesized scrambled images, we verified that the mean luminance and RMS contrast remained nearly identical to the original textures. The difference was within the floating-point error.

**Images with similar skewness and kurtosis.** We generated a new set of texture and scramble images in which we normalized the luminance histograms (scikit-image function ‘match\_histograms’). This function was applied before the rescaling procedure (see “Texture normalization”), therefore skewness and kurtosis values were not exactly equal between images after rescaling; however, by being randomly mixed, skewness and kurtosis could not be used to separate between textures families and textures from scrambles (Supplementary Fig. 3a). We then created a texture-scramble and texture-texture discrimination task in which mice used these new images ( $n = 7$  mice, newly trained). These mice had to discriminate between “scales” images and corresponding scrambles, and between “scales” and “plants” textures.

## Image analysis

**Image statistics.** We explored the image statistics at various levels of complexity. Our texture normalization procedure ensured that the pixel histogram distributions had identical means and standard deviations between the images (i.e., luminance, and RMS contrast). Within families and between a matching pair of textures and scrambles, we also confirmed that the average orientation and spatial frequency content were the same. To do so, in each image Fourier transform, we measured the average power in “slices” of the spatial frequencies and orientations (spatial frequency bins: [0.01 cpd, step 0.02 cpd: 0.5cpd]; orientation bins: [0°: step 15°: 180°]). The plots in Supplementary Fig. 1b, c show the amplitude values as a function of spatial frequencies and orientations, averaged across 20 exemplars for all families and stimulus types, and normalized to 1. To measure the higher-order statistics of the images, we decomposed them using an approach devised by Portilla and Simoncelli<sup>15</sup>, which decomposes an image using a bank of linear and energy filters tuned to different orientations, spatial frequencies, and spatial positions. The correlations are then computed across the outputs of these filters (i.e., the “PS statistics”). The parameters and classification of the PS statistics we adopted follow what has been previously described<sup>34,36,38,39</sup>. Briefly, we used a filter bank composed of four spatial scales (four downscaling octaves), four orientations (0°, 45°, 90°, 135°), and a spatial neighborhood of seven pixels to compute the filter output correlations. In addition, the marginal statistics of the pixel distributions were also computed (min, max, mean, standard deviation, skewness, and kurtosis). However, since part of our image synthesis pipeline procedure already ensured equal mean and standard deviation, only the differences in skewness and kurtosis were added to the characterization of the image statistics. In the end, the output of this image decomposition yielded four main groups of PS statistics: 1) marginal statistics (skewness and kurtosis); 2) spectral statistics; 3) linear cross-correlation statistics; and 4) energy cross-correlation statistics.

**Dimensionality reduction of PS statistics.** The number of parameters associated with the PS decomposition is larger (740) than the total number of images (320, eight image categories—four texture families and four scrambles—and 20 exemplars per category with two rotations). We thus reduced the number of parameters by applying Principal Component Analysis (PCA) to each PS statistical group after z-scoring the parameter values. We retained at most eight components in each group, which explained at least 70% of the variance per group (Supplementary Fig. 8). The marginal statistics with only two “dimensions” were excluded from this decomposition. After PCA, we again z-scored the outputs across exemplars to ensure that the range of parameter values between the groups of statistics was commensurate; this was necessary to gain interpretability of the distance metric later introduced, which was based on these reduced PS statistics. We also confirmed that the reduced PS statistics retained sufficient information to discriminate between textures and scrambles, with the energy cross-correlation statistics maximally distinguishing between them (Supplementary Fig. 9a,b).

## Imaging experiments

**Visual stimuli.** The visual stimuli were shown on a gamma-corrected monitor (widefield: Iiyama Prolite LE4041UHS 40”, two-photon: Iiyama Prolite B2776HDS-B1 27”). Except for the experiments for the discriminability gradient (see related section below), the size of the stimuli was 100° of visual angle with a raised cosine window for vignetting to correct for sharp edges; the stimuli were shown in front of the mouse perpendicular to its midline, which pointed to the center of the screen. The animal was at a distance of ~33 cm from the monitor for widefield experiments and ~24 cm for two-photon experiments. For widefield recordings, the stimuli were presented for 250 ms, followed by 750 ms of an isoluminant gray screen (ITI) before a new trial started.

Each mouse was shown 20 exemplars of four texture families and four scramble families (computed from the textures), a total of 10 times each exemplar, with 200 blank trials (i.e., trials with an isoluminant gray screen and no stimuli). This resulted in a total of 1600 trials with images and 200 trials with no stimulus (blanks). The presentation of each image/blank was fully randomized across the entire session.

The two-photon experiments followed the same temporal structure as the widefield experiments; however, we reduced the number of repeats and added image rotations. Specifically, each mouse was shown 20 exemplars of four texture families and four scramble families: a total of eight times for each exemplar, and two rotations (0° and 90°) of each exemplar, with 160 blank trials. This resulted in a total of 2560 trials with images and 160 trials with no stimuli (blanks). We also recorded the responses to oriented gratings: 100 degrees in size, four orientations (0°, 45°, 90°, 135°), five spatial frequencies (0.02, 0.04, 0.1, 0.2, 0.5 cpd) and 15 repeats per stimulus.

**Widefield imaging.** As described in ref. 103, awake mice were head-fixed and placed under a dual cube THT microscope (Brainvision Inc.) for widefield imaging in tandem-lens epifluorescence configuration using two AF NIKKOR 50 mm f/1.4D lenses. We imaged the jGCaMP7f fluorescence signals using interleaved shutter-controlled blue and violet LEDs with a CMOS camera (PCO Edge 5.5) with an acquisition framerate of 60 Hz. This dual color recording method ensured that we could capture both the calcium-dependent GCaMP signal (blue light path) as well as the hemodynamic-dependent signal (violet light path), as previously reported in other studies<sup>69</sup>. The blue light path consisted of a 465 nm centered LED (LEX-2, Brainvision Inc.), a 475 nm bandpass filter (Edmund Optics BP 475 × 25 nm OD4  $\phi$  = 50 mm), and two dichroic mirrors with 506 and 458 nm cutoff frequencies, respectively (Semrock FF506-Di03 50 × 70 mm, FF458-DFi02 50 × 70 mm). The violet path consisted of a 405 nm centered LED (Thorlabs M405L2 and LEDD1B driver), a 425 nm bandpass filter (Edmund Optics BP 425 × 25 nm OD4  $\phi$  = 25 mm), a collimator (Thorlabs COP5-A), and joined the blue LED path at the second dichroic mirror. The fluorescence light path traveled through the two dichroic mirrors (458 and 506 nm, respectively) and a 525 nm bandpass filter (Edmund Optics, BP 525 × 25 nm OD4  $\phi$  = 50 mm) and was finally captured with the PCO Edge 5.5 CMOS camera using the cameralink interface. Camera acquisition was synchronized to the LED illumination via a custom Arduino-controlled software. The frame exposure lasted 12 ms, starting 2 ms after opening each LED shutter to allow the LED illumination to stabilize. In a subset of the widefield experiments we displayed texture and scramble stimuli on a large screen, IYAMA Prolite LE4041UHS 40" monitor. Mice were placed 22 cm away from the monitor, with the body midline pointing at the right edge of the monitor. With these parameters the visual stimuli subtended approximately an azimuth range of [−62.4°, +62.4°] and an elevation range of [−48.5°, +48.5°].

**Preprocessing the widefield data.** Data preprocessing was done with custom Python and MATLAB code, with subsequent analyses done in Python. The continuously acquired imaging data were split into blue and violet channels. Then, as described in refs. 103,104, we corrected for the “hemodynamic component” by removing a calcium-independent component from the recorded signal. For every pixel, the blue and violet data were independently transformed into a relative fluorescence signal,  $\frac{\Delta F}{F} = (F - aF - b)/b$ , where  $F$  is the original data, and the  $a$  and  $b$  coefficients are obtained by linear fitting each time series, i.e.,  $F(t) \sim at - b$ . Afterwards, for each pixel, the violet  $\frac{\Delta F}{F}$  signal was low-pass filtered (6th order IIR filter with cutoff at 5 Hz) and linearly fitted to the blue  $\frac{\Delta F}{F}$  signal: the hemodynamic-corrected  $\frac{\Delta F}{F}$  signal was obtained as  $\frac{\Delta F}{F} \text{ corr} = \frac{\Delta F}{F} \text{ blue} - (c \frac{\Delta F}{F} \text{ violet} + d)$ , where  $c$  and  $d$  are the coefficients from linearly fitting the low-pass filtered  $\frac{\Delta F}{F} \text{ violet}$  to the  $\frac{\Delta F}{F} \text{ blue}$  signal, i.e.,  $\frac{\Delta F}{F} \text{ blue}(t) \sim c \frac{\Delta F}{F} \text{ violet}(t) - d$ . The continuously acquired data was then split into trial periods comprising

sequences of frames in a temporal window of [−500, +1000] ms relative to stimulus onset. This resulted in a tensor with seven dimensions: [stimulus type (texture or scramble), family type (4), exemplars (20), repeats (10), no. pixels X (256), no. pixels Y (230), no. frames]. Next, we averaged across repeats to obtain an “exemplar response tensor.”

**Retinotopy maps.** After the mice recovered from the cranial-window surgery (typically 3 to 4 days), we performed widefield imaging recordings during visual stimulation with counterphase flickering bars to obtain maps of retinotopy. We used a standard frequency-based method<sup>105</sup> with slowly moving horizontal and vertical flickering bars and corrections for spherical projections<sup>74</sup>. Visual area segmentation was performed based on azimuth and elevation gradient inversions<sup>73</sup>. The retinotopic maps were derived under light anesthesia (Isoflurane) with the animal midline pointing to the right edge of the monitor (IYAMA Prolite LE4041UHS 40"), centered relative to the monitor height, and with the animal's left eye at -25 cm from the center of the screen.

**Two-photon imaging.** As described in ref. 64, imaging experiments were performed using the two-photon imaging mode of the multiphoton confocal microscope (Model AIRMP, Nikon, Japan) with a Ti:sapphire laser (80 MHz, Coherent©, Chameleon Vision II). The microscope was controlled using the A1 software (Nikon). The objective was a 16x water immersion lens (NA, 0.8; working distance, 3 mm; Nikon). The field of view (512 × 512 pixels) was 532  $\mu\text{m}$  × 532  $\mu\text{m}$ . jGCaMP7f was excited at 920 nm, and the laser power was ~40 mW. Images were acquired continuously at a 30 Hz frame rate using a resonant scanner. To align the two-photon field of view with the maps of retinotopy, we captured a vascular image at the surface of the cortex and used it for reference.

**Preprocessing of two-photon data.** All the analyses, except for neuronal segmentation, were conducted using a custom code written in Python. Cells were segmented using Suite2p<sup>106</sup>, followed by the manual classification of the segmented ROIs. We then computed the  $\Delta F/F$  response values (%) for each neuron by first applying a neuropil correction:  $F_c = F_s - 0.7 \times F_n$ , where  $F_c$  is the corrected signal,  $F_s$  is the soma fluorescence, and  $F_n$  is the neuropil fluorescence. Then, we computed a baseline-fluorescence value ( $F_b$ ) as the mean of  $F_c$  during the first five seconds of the recordings when no stimuli were shown on the screen. We then detrended  $F_c$  (Scipy function `scipy.signal.detrend`) to remove the slow decrease in fluorescence sometimes observed across several tens of minutes and used the zero-mean detrended signal  $F_d$  to compute  $\Delta F/F = F_d/F_b$ .

#### Data analysis: widefield

**Defining regions of interest.** For every visual area, we defined a visually responsive ROI (or stimulus ROI) based on the maps of azimuth and elevation obtained from widefield imaging, so as to include a range of [+30°, −10°] in azimuth (relative to the contra- and ipsilateral visual fields, respectively) and elevation ( $\pm 30^\circ$ ), which, for the azimuth, was a conservative estimate of the retinotopic representation of the stimuli (of size  $\pm 50^\circ$  in azimuth and elevation).

**Peak-response and p-value maps.** Widefield responses to textures and scrambles (Fig. 2b) were computed by averaging across repeats, exemplars, and families; the frames were then averaged in a time widow [200, 400] ms after the stimulus onset, approximately centered around the time of peak response. The temporal response curves in V1 and LM to the textures and scrambles (Fig. 2c) were computed by averaging across repeats, families, and pixels within the response ROIs in V1 and LM; the variability was across the exemplars. The response ROIs were defined based on retinotopy as the cortical region that “mapped” the stimulus location in the visual space. The error bands

indicated a 95% confidence interval across the exemplars. To evaluate the significance of the differential response to the textures and scrambles (Fig. 2f), we tested against a distribution of pre-stimulus responses. Specifically, we first computed the response–difference distributions by subtracting the responses to texture exemplars (averaged across repeats) from the randomly paired scramble exemplars. As before, the frames were also averaged around the time of the peak response, [200, 400] ms after the stimulus onset. This resulted in a tensor with four dimensions: [family type (4), exemplars (20), no. pixels X (256), no. pixels Y (230), no. frames]. By grouping the responses to all the families and exemplars, we generated response–difference distributions for each pixel, each containing 80 data points. We applied the same procedure in a temporal window [−350, −100] ms prior to stimulus onset to obtain “null” distributions for each pixel. Finally, we tested for statistical differences between the pre- and post-stimulus onset distributions using a paired t-test and reporting the associated p-values. This procedure was applied to each animal, and the p-value maps were then used to compute the texture modulation of each visual area.

**Texture-scramble discriminability gradient in V1.** We used full field stimuli (azimuth [−62.4°, +62.4°], elevation [−48.5°, +48.5°]), sufficiently large to activate the entire V1. For each V1 camera pixel, we averaged responses across repeats, and time frames within a window of [200, 400] ms after stimulus onset. Then, we compared average discriminability values ( $d'$ , textures vs. scrambles) in the upper visual field ([0°, +40°] elevation, [−20°, +20°] in azimuth) vs. the lower visual field ([−40°, 0°] elevation, [−20°, +20°] in azimuth), and tested for a significant difference in average  $d'$  values using a t-test. Similarly, we measured the average discriminability values in the left visual field ([−40°, 0°] in azimuth, [−20°, +20°] elevation) vs. the right visual field ([0°, +40°] in azimuth, [−20°, +20°] elevation). Statistics were computed across 8 recording sessions ( $n = 4$  mice, 2 sessions each).

**Texture selectivity of visual areas.** To determine how significantly a visual area was modulated by textures compared to scrambles, we computed the proportion of the significantly modulated pixels ( $p < 0.01$ , from the p-value maps) within the stimulus ROI of each area (described in the section “Defining regions of interest”). This was separately computed in five visual areas (V1, LM, RL, AM, and PM) that were reliably segmented in all animals (Fig. 2g).

**Texture discriminability.** To compute the texture–scramble discriminability values for V1 and LM (Fig. 2h), we considered the responses to exemplars—separately for textures and scrambles—averaged across (i) repeats, (ii) pixels within stimulus ROIs (see section “Defining regions of interest”), and (iii) time frames within a window of [200, 400] ms after the stimulus onset. We then calculated a texture–scramble discriminability index ( $d'$ ) as follows:

$$d' = \frac{\mu_{tex} - \mu_{sc}}{\sqrt{\frac{1}{2}(\sigma_{tex}^2 + \sigma_{sc}^2)}} \quad (2)$$

Where  $\mu_{tex}$  and  $\mu_{sc}$  are the mean responses to the texture and scramble exemplars (80), and  $\sigma_{tex}^2$  and  $\sigma_{sc}^2$  the corresponding variances. To calculate the “null distribution” of the  $d'$  values shown in Fig. 2h (gray band), we followed the same procedure as above in a time window [−300, −0] ms prior to the stimulus onset, reporting the 5% and 95% percentiles of that distribution.

### Data analysis: two-photon

**Stimulus-responsive cells.** In a typical experiment, we could segment ~200–450 cells (as described in the section “Two-photon imaging”). To establish whether a cell was visually responsive, in each trial (−500, +1000] ms relative to stimulus onset) we “frame-zero” corrected  $\Delta F/F$

by subtracting the average activity within a pre-stimulus period of [−500, 0] ms. Then, we used a  $d'$  discriminability measure (similar to refs. 74,107) by comparing the responses to visual stimuli and to “blanks.” Specifically, in each trial and for every segmented cell, we averaged the responses in a window of [250, 500] ms post stimulus onset. We then used these average values to generate two distributions: one from the trials with visual stimuli, the other from the “blank” trials. The distributions with the visual stimuli were computed separately for the individual texture and scramble exemplars and considering the response variability across repeats. For each stimulus exemplar, we then computed a discriminability measure,  $d'_{stim}$ , as done in refs. 74,107:

$$d'_{stim} = \frac{\mu_{stim} - \mu_{blank}}{\sigma_{stim} + \sigma_{blank}} \quad (3)$$

Where,  $\mu_{stim}$  is the mean response across repeats for the chosen exemplar,  $\mu_{blank}$  is the mean response across the repeats of blank trials, and  $\sigma_{stim}$ ,  $\sigma_{blank}$  are the corresponding standard deviations. This procedure generated a distribution of  $d'_{stim}$  values for each cell. A cell was considered visually responsive if the maximum value of this distribution was  $\geq 1$ , and if  $\Delta F/F \geq 6\%$  in the stimulus-response window (for consistency with refs. 74,107). Subsequent analyses were performed on this subset of stimulus-responsive cells.

**Texture–scramble d-prime.** For every stimulus-responding cell, we considered frame-zero corrected  $\Delta F/F$  data, averaging across repeats and responses in a time window of [250, 500] ms after stimulus onset. We then considered the data variability across exemplars (and their rotations) to compute a discriminability measure  $d'$  as follows:

$$d' = \frac{\mu_{tex} - \mu_{sc}}{\sqrt{\frac{1}{2}(\sigma_{tex}^2 + \sigma_{sc}^2)}} \quad (4)$$

Where  $\mu_{tex}$  and  $\mu_{sc}$  are the mean responses to the texture and scramble exemplars, and  $\sigma_{tex}^2$  and  $\sigma_{sc}^2$  the corresponding variances.

**Regressive model.** Using a set of reduced PS statistics as regressors (see section “Image statistics”), we constructed a linear regressive model (ridge regularized) to predict individual cell responses. For each exemplar, we computed an average response value as the mean  $\Delta F/F$  (averaged across repeats and frame-zero corrected) in a time window of [250, 500] ms post stimulus onset. For each neuron  $i$ , the model was trained to capture the responses to different exemplars using the following loss function:

$$\min_{w_i} \|y_i - Xw_i\|_2^2 + \lambda \|w_i\|_2^2 \quad (5)$$

Where  $w_i$  are the optimization weights,  $y_i$  the data,  $\lambda$  a regularization parameter, and  $X$  the reduced PS statistics (two dimensions per group, i.e., the first two PCs). We confirmed that the model did not perform significantly better when using more PCs. The model was trained with five-fold cross validation to reduce overfitting, and the regularization parameter  $\lambda$  was optimized using a grid search. The model's performance was evaluated in terms of the explained variance (EV) in the cross-validated data. To establish the significance of the model's fit and to derive an EV threshold value for the inclusion of cells in the analyses of Fig. 3l, we used a permutation test. For a given cell, we refitted the responses using as input statistics those from randomly chosen images (across exemplars from all textures and scrambles). Therefore, for each experiment, we obtained a shuffled distribution of EVs (across cells) and chose the 95<sup>th</sup> percentile of the distribution as the threshold value for significance ( $\alpha = 0.05$ ). We used this approach in all  $n = 20$  experiments, resulting in an average threshold value,

$EV_{th} = 0.87\% \pm 0.07\%$  (s.e.). We set a conservative inclusion threshold at  $EV_{th} = 1\%$ .

**Regressive model: weight analysis.** To examine the contribution of the different reduced PS statistics in the regressive model, we summed the absolute values of the regressive weights separately for each of the four statistical groups: for a given cell, and for the PS group  $i$ , we computed  $W_i = \sum_{j=1}^d |w_{ij}|$ , with  $d = 2$ , that is, the number of PCs for the reduced PS statistics. We then averaged  $W_i$  across all the cells in a given animal (individual data points in Fig. 3l).

**Regressive model: unique EV.** To examine the unique contribution to the explained variance by the different reduced PS statistics, we measured the loss in EV when training models without a particular statistical group. Specifically, considering a subset of cells with significant explained variance ( $EV > 10\%$ ), we first trained a model with all four groups of PS statistics (full model). Further, we trained four more models, each missing one of the four PS groups. We then computed a measure of unique variance explained,  $\Delta EV_{u_i}$ , as follows:

$$\Delta EV_{u_i} = 100 \frac{EV_f - EV_i}{EV_f} \forall i \in \{PS_1, \dots, PS_4\}. \quad (6)$$

Where  $EV_i$  is the explained variance of a model trained without the PS group  $i$ , and  $EV_f$  is the explained variance of the full model.

**PCA embedding of neural responses.** For every stimulus-responding cell, we considered the frame-zero corrected  $\Delta F/F$  data, averaging across repeats and time frames in a time window of [250, 500] ms post stimulus onset. After z-scoring the responses of each cell to different exemplars, we applied PCA ( $n = 20$  PCs, separately for V1 and LM populations) to “standardize” the population size, thus facilitating a comparison between experiments, each having a different number of segmented cells. An example of a PCA space of neural activity is shown in Fig. 4g for LM recordings ( $n = 2$  PCs).

**Decoding responses to textures and scrambles.** In the PCA spaces of neural activity for V1 and LM, as described in the section above, we considered responses to exemplars separately for each of the four texture–scramble families. For each family, we trained a binary logistic classifier to distinguish texture exemplars from scramble exemplars. The model was five-fold cross-validated, and its performance was evaluated using the average accuracy across the five folds. We repeated the same analysis by varying the number of PCs and examining the related changes in classification accuracy separately for the V1 and LM data (Supplementary Fig. 11a–c). Instead, for the analysis in Supplementary Fig. 9a, a binary classifier was trained to discriminate between texture and scramble images (across all families and exemplars), separately on different PS statistical groups.

**Distance metrics for stimulus statistics.** For each of the four PS statistical groups, we considered a 2D-PCA space of image statistics (see section “PCA of PS statistics”), with two PCs already sufficient for near-optimal classification performance (Supplementary Fig. 9a). The overall distance patterns described in Fig. 4 were consistent when using larger numbers of PCs. A single point in each PCA space corresponds to the statistical representation of an exemplar image based on the associated PS statistical decomposition (reduced to four main PS statistical groups). To compute the radius of a cloud of points (20 exemplar points) for a given family, we computed the standard deviation of the  $x$  and  $y$  coordinates,  $\sigma_x, \sigma_y$ , and defined the radius as their mean value  $r_i = \frac{\sigma_x + \sigma_y}{2}$ . For the inter-cluster distance of a given pair of clouds (i.e., exemplars of textures or scrambles), we first computed the center of mass of the two clouds as the mean of the  $x$  and  $y$  coordinates,  $\mu_x, \mu_y$ , measured their Euclidean distance which we then

normalized (divided) by the mean of the two corresponding radii. The inter-cluster distances were calculated for all the matching pairs of texture/scramble families (for Fig. 4c). Further, the radius values were computed for all the families and stimulus types and for all the groups of PS statistics.

**Decoding the responses to texture families.** In the PCA spaces of neural activity for V1 and LM (as described in the section “PCA embedding of neural responses”), we created a linear decoding model trained to classify all four texture families. We used a multinomial logistic classifier with an L1 regularization penalty. The training data consisted of the cells’ responses to 160 texture stimuli (four families, 20 exemplars, two rotations). The model was trained using five-fold cross validation, and the regularization factor was optimized with a grid search. The model’s performance was evaluated as the cross-validated accuracy averaged across folds. We also examined the dependence of the model’s performance on the number of PCs (Fig. 4h).

**Distance metrics for neural representations.** To compare the representational differences between V1 and LM, we created a common PCA space of neuronal activations. For a given mouse, we considered responses to exemplars pre-processed as described in “PCA embedding of neural responses” (before PCA). We then applied PCA to a “concatenated” ensemble of V1 and LM cells to derive a common PCA space with  $n = 16$  components. The number of segmented cells and the z-scored response values were commensurate between V1 and LM. Using the PCA projection matrix, and by zeroing responses of the “other” area, we could then separately project the V1 and LM responses in this common space. We then measured the radii of the activation “clouds” in this PCA space for each texture family, as well as the inter-cluster distances for pairs of texture families. To compute the radius of a cloud of points ( $n = 40$  points, 20 exemplars, 2 rotations) for a given family, we computed the standard deviation of the  $x$  and  $y$  coordinates,  $\sigma_x, \sigma_y$ , and defined the radius as their mean value  $r_i = \frac{\sigma_x + \sigma_y}{2}$ . For the inter-cluster distance of a given pair of clouds (i.e., exemplars of two textures), we first computed the center of mass of the two clouds as the mean of the  $x$  and  $y$  coordinates,  $\mu_x, \mu_y$ , measured their Euclidean distance. Finally, we compared the radii and inter-cluster distances for all six pairs of families between V1 and LM.

For the Mahalanobis distance analysis, for each animal we considered the clouds of points in the V1-LM shared PCA space and computed a Mahalanobis distance value for each of the six pairs of texture families. An ANOVA statistical test was then performed to quantify the area effect (V1 vs. LM).

### Reporting summary

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

### Data availability

Data to generate all figure panels is provided as a Source Data file and deposited in <https://github.com/CBS-NCB/mouseTextures> Source data are provided with this paper.

### Code availability

Analysis code is available at this GitHub repository: <https://github.com/CBS-NCB/mouseTextures>

### References

- Victor, J. D., Conte, M. M. & Chubb, C. F. Textures as probes of visual processing. *Annu. Rev. Vis. Sci.* **3**, 275–296 (2017).
- Julesz, B. Textons, the elements of texture perception, and their interactions. *Nature* **290**, 91–97 (1981).

3. Todd, J. T., Norman, J. F., Koenderink, J. J. & Kappers, A. M. L. Effects of texture, illumination, and surface reflectance on stereoscopic shape perception. *Perception* **26**, 807–822 (1997).
4. Todd, J. T., Oomes, A. H. J., Koenderink, J. J. & Kappers, A. M. L. The perception of doubly curved surfaces from anisotropic textures. *Psychol. Sci.* **15**, 40–46 (2004).
5. Schmid, A. M. & Victor, J. D. Possible functions of contextual modulations and receptive field nonlinearities: Pop-out and texture segmentation. *Vision Res.* **104**, 57–67 (2014).
6. Li, A. & Zaidi, Q. Perception of three-dimensional shape from texture is based on patterns of oriented energy. *Vision Res.* **40**, 217–242 (2000).
7. Bergen, J. R. & Julesz, B. Parallel versus serial processing in rapid pattern discrimination. *Nature* **303**, 696–698 (1983).
8. Jagadeesh, A. V. & Gardner, J. Texture-like representation of objects in human visual cortex. *Proc. Natl Acad. Sci.* **119**, e2115302119 (2022).
9. Srivastava, A., Lee, A. B., Simoncelli, E. P. & Zhu, S. C. On advances in statistical modeling of natural images. *J. Math. Imaging Vis.* **18**, 17–33 (2003).
10. Torralba, A. & Oliva, A. Statistics of natural image categories. *Netw.: Comput. Neural Syst.* **14**, 391 (2003).
11. Ruderman, D. The statistics of natural images. *Netw.: Comput. Neural Syst.* **5**, 517 (1994).
12. Leung, T. & Malik, J. Representing and recognizing the visual appearance of materials using three-dimensional textons. *Int. J. Comput. Vis.* **43**, 29–44 (2001).
13. Victor, J. D. & Conte, M. M. Local image statistics: maximum-entropy constructions and perceptual salience. *J. Opt. Soc. Am. A* **29**, 1313–1345 (2012).
14. Chubb, C., Econopoulou, J. & Landy, M. S. Histogram contrast analysis and the visual segregation of IID textures. *J. Opt. Soc. Am. A* **11**, 2350–2374 (1994).
15. Portilla, J. & Simoncelli, E. P. A parametric texture model based on joint statistics of complex wavelet coefficients. *Int. J. Comput. Vis.* **40**, 49–70 (2000).
16. Bar-Joseph, Z., El-Yaniv, R., Lischinski, D. & Werman, M. Texture mixing and texture movie synthesis using statistical learning. *IEEE Trans. Vis. Comput. Graph.* **7**, 120–134 (2001).
17. de Bonet, J. & Viola, P. A Non-parametric multi-scale statistical model for natural images. *NIPS* <https://proceedings.neurips.cc/paper/1997/hash/c5cc17e395d3049b03e0f1cceb02b4d-Abstract.html> (1997).
18. Balas, B. J. Texture synthesis and perception: Using computational models to study texture representations in the human visual system. *Vision Res.* **46**, 299–309 (2006).
19. Gatys, L. A., Ecker, A. S. & Bethge, M. Texture synthesis using convolutional neural networks. *Adv. Neural Inf. Process Syst.* **2015**, 262–270 (2015).
20. Gatys, L. A., Ecker, A. S. & Bethge, M. Texture and art with deep neural networks. *Curr. Opin. Neurobiol.* **46**, 178–186 (2017).
21. Vacher, J., Dvaila, A., Kohn, A. & Coen-Cagli, R. Texture interpolation for probing visual perception. In *NeurIPS*, 22146–22157 <https://dl.acm.org/doi/abs/10.5555/3495724.3497581> (2020).
22. Barlow, H. B. Possible principles underlying the transformation of sensory messages. *Sens. Commun.* **1**, 217–234 (1961).
23. Sterling, P. & Laughlin, S. Principles of neural design. *The MIT Press* <https://doi.org/10.7551/MITPRESS/9780262028707.001.0001> (2015).
24. Atick, J. J. & Redlich, A. N. Towards a theory of early visual processing. *Neural Comput.* **2**, 308–320 (1990).
25. Laughlin, S. A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch C Biosci.* **36**, 910–912 (1981).
26. Olshausen, B. A. & Field, D. J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381**, 607–609 (1996).
27. Pitkow, X. & Meister, M. Decorrelation and efficient coding by retinal ganglion cells. *Nat. Neurosci.* **15**, 628–635 (2012).
28. Zador, A. M. A critique of pure learning and what artificial neural networks can learn from animal brains. *Nat. Commun.* **10**, 1–7 (2019).
29. Caramellino, R. et al. Rat sensitivity to multipoint statistics is predicted by efficient coding of natural scenes. *eLife* **10**, e72081 (2021).
30. Merigan, W. H. Cortical area V4 is critical for certain texture discriminations, but this effect is not dependent on attention. *Vis. Neurosci.* **17**, 949–958 (2000).
31. Hanazawa, A. & Komatsu, H. Influence of the direction of elemental luminance gradients on the responses of V4 cells to textured surfaces. *J. Neurosci.* **21**, 4490–4497 (2001).
32. Arcizet, F., Jouffrais, C. & Girard, P. Natural textures classification in area V4 of the macaque monkey. *Exp. Brain Res.* **189**, 109–120 (2008).
33. Nandy, A. S., Sharpee, T. O., Reynolds, J. H. & Mitchell, J. F. The fine structure of shape tuning in Area V4. *Neuron* **78**, 1102–1115 (2013).
34. Freeman, J., Ziemba, C. M., Heeger, D. J., Simoncelli, E. P. & Movshon, J. A. A functional and perceptual signature of the second visual area in primates. *Nat. Neurosci.* **16**, 974–981 (2013).
35. Ziemba, C. M., Freeman, J., Simoncelli, E. P. & Movshon, J. A. Contextual modulation of sensitivity to naturalistic image structure in macaque V2. *J. Neurophysiol.* (2018) <https://doi.org/10.1152/jn.00900.2017>.
36. Okazawa, G., Tajima, S. & Komatsu, H. Image statistics underlying natural texture selectivity of neurons in macaque V4. *Proc. Natl Acad. Sci. USA* **112**, E351–V60 (2015).
37. Yu, Y., Schmid, A. M. & Victor, J. D. Visual processing of informative multipoint correlations arises primarily in V2. *eLife* **2015**, 1–13 (2015).
38. Ziemba, C. M., Freeman, J., Movshon, J. A. & Simoncelli, E. P. Selectivity and tolerance for visual texture in macaque V2. *Proc. Natl Acad. Sci. USA* **113**, 3140–3149 (2016).
39. Okazawa, G., Tajima, S. & Komatsu, H. Gradual development of visual texture-selective properties between Macaque Areas V2 and V4. *Cerebr. Cortex* **27**, 4867–4880 (2017).
40. Epstein, R. A. & Baker, C. I. Scene perception in the human brain. *Annu. Rev. Vis. Sci.* **5**, 373–397, <https://doi.org/10.1146/annurev-ision-091718-014809> (2019).
41. Rust, N. C. & DiCarlo, J. J. Selectivity and Tolerance (“Invariance”) both increase as visual information propagates from cortical area V4 to IT. *J. Neurosci.* **30**, 12978–12995 (2010).
42. Purpura, K. P., Victor, J. D. & Katz, E. Striate cortex extracts higher-order spatial correlations from visual textures. *Proc. Natl Acad. Sci. USA* **91**, 8482–8486 (1994).
43. Freeman, J. & Simoncelli, E. P. Metamers of the ventral stream. *Nat. Neurosci.* (2011) <https://doi.org/10.1038/nn.2889nn.2889>.
44. Parthasarathy, N. & Simoncelli, E. P. Self-supervised learning of a biologically-inspired visual texture model. *arXiv preprint arXiv:2006.16976* <https://doi.org/10.48550/arXiv.2006.16976> (2020).
45. Laskar, M. N. U., Giraldo, L. G. S. & Schwartz, O. Deep neural networks capture texture sensitivity in V2. *J. Vis.* **20**, 21–21 (2020).
46. Bergen, J. R. & Adelson, E. H. Early vision and texture perception. *Nature* **333**, 363–364 (1988).
47. Malik, J. & Perona, P. Preattentive texture discrimination with early vision mechanisms. *J. Opt. Soc. Am. A* **7**, 923–932 (1990).
48. Felsen, G., Touryan, J., Han, F. & Dan, Y. Cortical sensitivity to visual features in natural scenes. *PLoS Biol.* **3**, e342 (2005).

49. Niell, C. M. & Scanziani, M. How Cortical circuits implement cortical computations: mouse visual cortex as a model. *Annu. Rev. Neurosci.* **44**, 517–546 (2021).
50. Zoccolan, D., Cox, D. D. & Benucci, A. Editorial: What can simple brains teach us about how vision works. *Front. Neural Circuits* **9**, 51 (2015).
51. Coogan, T. A. & Burkhalter, A. Hierarchical organization of areas in rat visual cortex. *J. Neurosci.* **13**, 3749–3772 (1993).
52. Wang, Q. & Burkhalter, A. Area map of mouse visual cortex. *J. Comp. Neurol.* **502**, 339–357 (2007).
53. Wang, Q., Gao, E. & Burkhalter, A. Gateways of ventral and dorsal streams in mouse visual cortex. *J. Neurosci.* **31**, 1905–1918 (2011).
54. D’Souza, R. D. et al. Hierarchical and nonhierarchical features of the mouse visual cortical network. *Nat. Commun.* **13**, 1–14 (2022).
55. Tafazoli, S. et al. Emergence of transformation-tolerant representations of visual objects in rat lateral extrastriate cortex. *eLife* **6**, e22794 (2017).
56. Matteucci, G., Marotti, R. B., Riggi, M., Rosselli, F. B. & Zoccolan, D. Nonlinear processing of shape information in rat lateral extrastriate cortex. *J. Neurosci.* **39**, 1649–1670 (2019).
57. Froudarakis, E. et al. Object manifold geometry across the mouse cortical visual hierarchy. *bioRxiv* 2020.08.20.258798 (2020) <https://doi.org/10.1101/2020.08.20.258798>.
58. Yu, Y., Stirman, J. N., Dorsett, C. R. & Smith, S. L. Selective representations of texture and motion in mouse higher visual areas. *Curr. Biol.* (2022) <https://doi.org/10.1016/J.CUB.2022.04.091>.
59. Ding, K., Ma, K., Wang, S. & Simoncelli, E. P. Image quality assessment: unifying structure and texture similarity. *ArXiv* (2020) <https://doi.org/10.1109/TPAMI.2020.3045810>.
60. Chubb, C. & Nam, J. H. Variance of high contrast textures is sensed using negative half-wave rectification. *Vision Res.* **40**, 1677–1694 (2000).
61. Chubb, C., Landy, M. S. & Econopoulou, J. A visual mechanism tuned to black. *Vision Res.* **44**, 3223–3232 (2004).
62. Thomson, M. G. A., Foster, D. H. & Summers, R. J. Human sensitivity to phase perturbations in natural images: A statistical framework. *Perception* **29**, 1057–1069 (2000).
63. Malach, R. et al. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc. Natl Acad. Sci. USA* **92**, 8135–8139 (1995).
64. Aoki, R., Tsubota, T., Goya, Y. & Benucci, A. An automated platform for high-throughput mouse behavior and physiology with voluntary head-fixation. *Nat. Commun.* **8**, 1196 (2017).
65. Lyamzin, D. R., Aoki, R., Abdolrahmani, M. & Benucci, A. Probabilistic discrimination of relative stimulus features in mice. *Proc. Natl Acad. Sci.* **118**, e2103952118 (2021).
66. Berditchevskaia, A., Cazé, R. D. & Schultz, S. R. Performance in a GO/NOGO perceptual task reflects a balance between impulsive and instrumental components of behaviour. *Sci. Rep.* **6**, 1–15 (2016).
67. Luongo, F. J. et al. Mice and primates use distinct strategies for visual segmentation. *eLife* **12**, e74394 (2023).
68. Gardner, J. L. Optimality and heuristics in perceptual neuroscience. *Nat. Neurosci.* **22**, 514–523 (2019).
69. Couto, J. et al. Chronic, cortex-wide imaging of specific cell populations during behavior. *Nat. Protoc.* **16**, 3241–3263 (2021).
70. La Chioma, A., Bonhoeffer, T. & Hübener, M. Disparity sensitivity and binocular integration in mouse visual cortex areas. *J. Neurosci.* JN-RM-1060-20 (2020) <https://doi.org/10.1523/JNEUROSCI.1060-20.2020>.
71. Sit, K. K. & Goard, M. J. Distributed and retinotopically asymmetric processing of coherent motion in mouse visual cortex. *Nat. Commun.* **11**, 1–14 (2020).
72. Qiu, Y. et al. Natural environment statistics in the upper and lower visual field are reflected in mouse retinal specializations. *Curr. Biol.* **31**, 3233–3247 (2021).
73. Garrett, M. E., Nauhaus, I., Marshel, J. H. & Callaway, E. M. Topography and areal organization of mouse visual cortex. *J. Neurosci.* **34**, 12587–12600 (2014).
74. Marshel, J. H., Garrett, M. E., Nauhaus, I. & Callaway, E. M. Functional specialization of seven mouse visual cortical areas. *Neuron* **72**, 1040–1054 (2011).
75. Musall, S., Kaufman, M. T., Juavinett, A. L., Gluf, S. & Churchland, A. K. Single-trial neural dynamics are dominated by richly varied movements. *Nat. Neurosci.* **22**, 1677–1686 (2019).
76. Caelli, T. & Julesz, B. On perceptual analyzers underlying visual texture discrimination. *Biol. Cybern.* **28**, 167–175 (1978).
77. Landy, M. S. & Bergen, J. R. Texture segregation and orientation gradient. *Vision Res.* **31**, 679–691 (1991).
78. Balas, B. & Conlin, C. Invariant texture perception is harder with synthetic textures: Implications for models of texture processing. *Vision Res.* **115**, 271–279 (2015).
79. Jia, X. et al. Multi-regional module-based signal transmission in mouse visual cortex. *Neuron* **110**, 1585–1598 (2022).
80. Goldbach, H. C., Akitake, B., Leedy, C. E. & Histed, M. H. Performance in even a simple perceptual task depends on mouse secondary visual areas. *eLife* **10**, e62156 (2021).
81. Goltstein, P. M., Reinert, S., Bonhoeffer, T. & Hübener, M. Mouse visual cortex areas represent perceptual and semantic features of learned visual categories. *Nat. Neurosci.* **24**, 1441–1451 (2021).
82. Kumar, M. G., Hu, M., Ramanujan, A., Sur, M. & Murthy, H. A. Functional parcellation of mouse visual cortex using statistical techniques reveals response-dependent clustering of cortical processing areas. *PLoS Comput Biol* **17**, e1008548 (2021).
83. Jin, M. & Glickfeld, L. L. Mouse Higher Visual Areas Provide Both Distributed and Specialized Contributions to Visually Guided Behaviors. *Curr. Biol.* **30**, 4682–4692.e7 (2020).
84. de Vries, S. E. J. et al. A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. *Nat. Neurosci.* **23**, 138–151 (2020).
85. Wang, Q., Sporns, O. & Burkhalter, A. Network analysis of corticocortical connections reveals ventral and dorsal processing streams in mouse visual cortex. *J. Neurosci.* **32**, 4386–4399 (2012).
86. Glickfeld, L. L. & Olsen, S. R. Higher-Order Areas of the Mouse Visual Cortex. *Annu. Rev. Vis. Sci.* (2017) <https://doi.org/10.1146/annurev-vision-102016-061331>.
87. Chung, S. Y. & Abbott, L. F. Neural population geometry: An approach for understanding biological and artificial neural networks. *Curr. Opin. Neurobiol.* **70**, 137–144 (2021).
88. Kriegeskorte, N. & Wei, X.-X. Neural tuning and representational geometry. *Nat Rev Neurosci* **22**, 703–718 (2021).
89. Hung, C. P., Kreiman, G., Poggio, T. & DiCarlo, J. J. Fast readout of object identity from macaque inferior temporal cortex. *Science* (1979) **310**, 863–866 (2005).
90. DiCarlo, J. J. & Cox, D. D. Untangling invariant object recognition. *Trends Cogn Sci* **11**, 333–341 (2007).
91. Tenenbaum, J. B., de Silva, V. & Langford, J. C. A global geometric framework for nonlinear dimensionality reduction. *Science* (1979) **290**, 2319–2323 (2000).
92. Duncker, L. & Sahani, M. Dynamics on the manifold: Identifying computational dynamical activity from neural population recordings. *Curr. Opin. Neurobiol.* **70**, 163–170 (2021).
93. Benucci, A. Motor-related signals support localization invariance for stable visual perception. *PLoS Comput Biol* **18**, e1009928 (2022).
94. Angelucci, A. et al. Circuits for local and global signal integration in primary visual cortex. *J. Neurosci.* **22**, 8633–8646 (2002).



95. Okazawa, G., Tajima, S. & Komatsu, H. Gradual development of visual texture-selective properties between Macaque Areas V2 and V4. *Cerebr. Cortex* **27**, 4867–4880 (2016).
  96. Ress, D., Backus, B. T. & Heeger, D. J. Activity in primary visual cortex predicts performance in a visual detection task. *Nat. Neurosci.* **3**, 940–945 (2000).
  97. Joseph, J. S., Chun, M. M. & Nakayama, K. Attentional requirements in a ‘preattentive’ feature search task. *Nature* **387**, 805–807 (1997).
  98. Stanislaw, H. & Todorov, N. Calculation of signal detection theory measures. *Behav. Res. Methods Instrum. Comput.* **31**, 137–149 (1999).
  99. Burgess, C. P. et al. High-yield methods for accurate two-alternative visual psychophysics in head-fixed mice. *Cell Rep.* **20**, 2513–2524 (2017).
  100. Luongo, F. J. et al. Mice and primates use distinct strategies for visual segmentation. *bioRxiv* 2021.07.04.451059 (2021) <https://doi.org/10.1101/2021.07.04.451059>.
  101. Simonyan, K. & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* (2014) <https://doi.org/10.48550/arxiv.1409.1556>.
  102. Prusky, G. T. & Douglas, R. M. Characterization of mouse cortical spatial vision. in *Vision Res.* 44 3411–3418 (Elsevier Ltd, 2004).
  103. Orlandi, J. G., Abdolrahmani, M., Aoki, R., Lyamzin, D. R. & Benucci, A. Distributed context-dependent choice information in mouse posterior cortex. *Nat. Commun.* **14**, 1–16 (2023).
  104. Abdolrahmani, M., Lyamzin, D. R., Aoki, R. & Benucci, A. Attention separates sensory and motor signals in the mouse visual cortex. *Cell Rep.* **36**, 109377 (2021).
  105. Kalatsky, V. A. & Stryker, M. P. New paradigm for optical imaging: temporally encoded maps of intrinsic signal. *Neuron* **38**, 529–545 (2003).
  106. Pachitariu, M. et al. Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *bioRxiv* (2017) <https://doi.org/10.1101/061507>.
  107. Juavinett, A. L. & Callaway, E. M. Pattern and component motion responses in mouse visual cortical areas. *Curr. Biol.* **25**, 1759–1764 (2015).
- (to A.B.); RIKEN JRA, University of Tokyo IST-RA, and JSPS-DC2 (to F.B.); HFSP postdoctoral fellowship LT000582/2019 (to J.O.); and SPDR for R.A.

### Author contributions

A.B., J.G., F.B., and J.O. designed the study. A.V.J. facilitated image generation. F.B. created the behavioral protocol, collected all behavioral data, performed all recordings, and conducted all analyses. A.B. and F.B. wrote the manuscript. R.A. made helpful comments on the behavioral protocol.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-45919-3>.

**Correspondence** and requests for materials should be addressed to Javier G. Orlandi or Andrea Benucci.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

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

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

© The Author(s) 2024

### Acknowledgements

We thank D. Zoccolan for his feedback on the interpretation of our findings. We thank Yuki Goya, Rie Nishiyama, and Yuka Iwamoto for their support with behavioral training, animal care, and surgeries. We thank O’Hara and Co., Ltd., for their support with the equipment. This work was funded by RIKEN BSI and RIKEN CBS institutional funding, JSPS grants 26290011, 17H06037, and CO219129, and a Fujitsu collaborative grant