



# V1 neurons are tuned to perceptual borders in natural scenes

Paolo Papale<sup>a,b,1</sup> , Wietske Zuiderbaan<sup>c,d,1</sup>, Rob R. M. Teeuwen<sup>a</sup>, Amparo Gilhuis<sup>a</sup>, Matthew W. Self<sup>a,2</sup> , Pieter R. Roelfsema<sup>a,e,f,g,2,3</sup> , and Serge O. Dumoulin<sup>c,d,h,i,2</sup>

Affiliations are included on p. 9.

Edited by Marlene Behrmann, University of Pittsburgh, Pittsburgh, PA; received January 1, 2023; accepted September 30, 2024

The visual system needs to identify perceptually relevant borders to segment complex natural scenes. The primary visual cortex (V1) is thought to extract local borders, and higher visual areas are thought to identify the perceptually relevant borders between objects and the background. To test this conjecture, we used natural images that had been annotated by human observers who marked the perceptually relevant borders. We assessed the effect of perceptual relevance on V1 responses using human neuroimaging, macaque electrophysiology, and computational modeling. We report that perceptually relevant borders elicit stronger responses in the early visual cortex than irrelevant ones, even if simple features, such as contrast and the energy of oriented filters, are matched. Moreover, V1 neurons discriminate perceptually relevant borders surprisingly fast, during the early feedforward-driven activity at a latency of ~50 ms, indicating that they are tuned to the features that characterize them. We also revealed a delayed, contextual effect that enhances the V1 responses that are elicited by perceptually relevant borders at a longer latency. Our results reveal multiple mechanisms that allow V1 neurons to infer the layout of objects in natural images.

V1 | neurophysiology | fmri | nonhuman primates | natural vision

The visual scenes that we perceive are filled with objects. We readily identify the extent of the objects and their borders, a process that is important for our understanding of an image's meaning. A previous study developed the Berkeley Segmentation data Set (BSD), which is a library of natural images, and demonstrated that the judgments of people who are asked to segment regions occupied by objects and their borders are highly consistent (Fig. 1*A*) (1). Despite appearing effortless, the segmentation process requires subjects to discount the physical properties of the image, such as local contrast (Fig. 1*B* and *C*). Fig. 1*A* shows the most relevant borders of an example image of the BSD, which was segmented by human observers, and Fig. 1*B* shows two isolated image patches. Although both patches contain a contrast-defined border (Fig. 1*C*), only the object border is relevant to perception and was segmented by the observers. Thus, local contrast does not uniquely identify perceptually relevant borders. Nevertheless, boundary perception in natural images influences vision because image elements at object borders are better perceived than image elements at less relevant image locations (2, 3).

In a review in 2005, Olshausen and Fields (4) argued that models of simple and complex cells in the primary visual cortex (V1) based on oriented filters (5) are poorly equipped to detect the edges of objects. Newer approaches that use deep convolutional neural networks explain more variance of V1 responses (6). However, when applied to natural images, these models do not identify the perceptually relevant borders that are marked by human observers either (Fig. 1*D*), and their output only bears a weak relation to scene perception. One possible reason is that these filter and neural network models primarily use feedforward connections that propagate information from lower to higher network levels. Another possibility is that V1 is not sensitive to the perceptual relevance, even though previous neurophysiological studies, which used artificial stimuli, demonstrated that segmentation signals influence V1 activity: The responses of V1 neurons increase if their RF is centered on an elongated contour that extends well beyond their RF (7, 8) (Fig. 1*F*). These influences are expressed during a delayed phase of the neuronal responses and presumably rely on feedback from downstream visual regions where neurons are tuned to more complex features of artificial stimuli, such as textures and displays with many line elements (7–16) (Fig. 1*F*).

Here, we asked how perceptual borders influence neuronal representations in the early visual cortex (17). We used the BSD to study the representation of object borders with human neuroimaging, electrophysiology in both awake and anesthetized monkeys, and computational modeling. We compared the neuronal activity elicited by object borders and other image patches with similar contrast-defined borders (Fig. 1*C* and *D* and

## Significance

How do we visually segregate objects from other objects and the background? Humans readily see the boundaries that demarcate objects, but it is not well understood how boundaries are represented in the visual cortex. We used primate electrophysiology, human neuroimaging, and computational modeling to investigate how neurons in the visual brain detect and represent object boundaries and how they distinguish them from other image regions with a similar level of contrast or orientation energy. We report that the representation of object borders is enhanced in the early visual cortex. Enhanced activity during the start of neuronal responses reflects the tuning to features in the receptive field. A later response enhancement also reflects contextual influences, mediated by lateral and feedback connections.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2024 the Author(s). Published by PNAS. This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

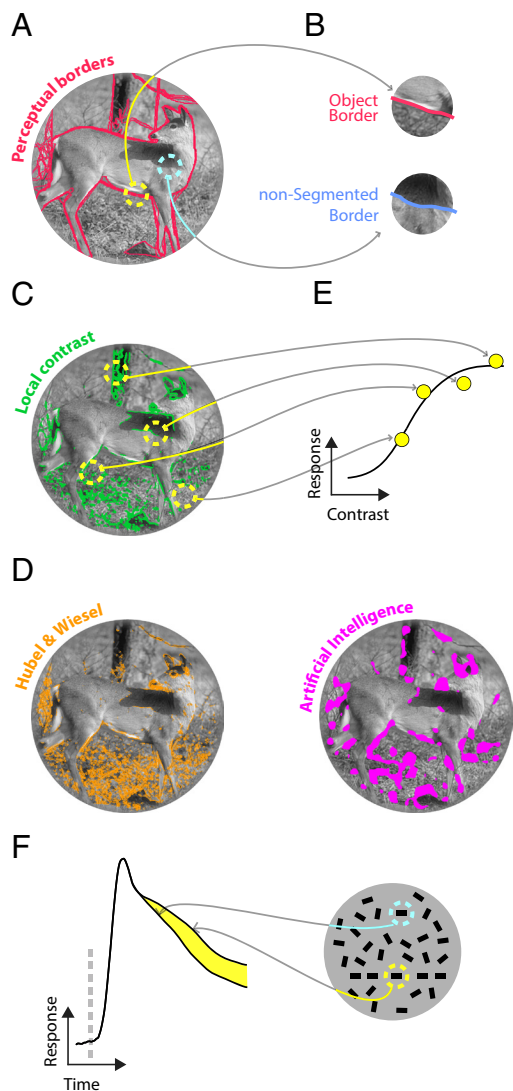
<sup>1</sup>P.P. and W.Z. contributed equally to this work.

<sup>2</sup>M.W.S., P.R.R., and S.O.D. contributed equally to this work.

<sup>3</sup>To whom correspondence may be addressed. Email: [p.roelfsema@nin.knaw.nl](mailto:p.roelfsema@nin.knaw.nl).

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2221623121/-/DCSupplemental>.

Published November 4, 2024.



**Fig. 1.** Identifying borders in natural scenes. (A) Perceptually relevant borders as segmented by human observers. The image comes from the BSD (1). (B) Two image patches that illustrate that local contrast does not predict perceptually relevant borders. (C) Pixels with the top 10% local image contrast are shown in green (1). (D) Pixels within the top 10% energy according to a complex-cell model (yellow, *Left*) and top 10% energy of the VGG-19 AI model, layer "conv3\_1" (*Right*, purple). Neither local contrast nor current V1 models capture the borders that are perceptually most relevant. (E) We computed the CRFs evoked in the visual cortex by image elements based on the local contrast or orientation energy in their (p)RF. (F) The responses of V1 neurons are enhanced when their RF falls on an elongated contour (yellow circle) in a stimulus with many line elements. This response modulation represents a recurrent influence from outside the neurons' RF.

*SI Appendix, Fig. S1*) in the early visual cortex. We report that V1 neurons exhibit a sensitivity to perceptually relevant object borders that goes beyond what is predicted by current models of feedforward V1 tuning.

## Results

**Stronger Responses to Object Borders in the Early Visual Cortex.** We investigated the activity elicited by perceptually relevant borders in natural images using human neuroimaging, multiunit activity in awake macaque monkeys, and single neurons in anesthetized monkeys (Fig. 2), using images from the BSD (1). The BSD allowed us to compare the neuronal responses elicited by contours of the same contrast that had been labeled by the observers as object contours or had not been labeled (*SI Appendix,*

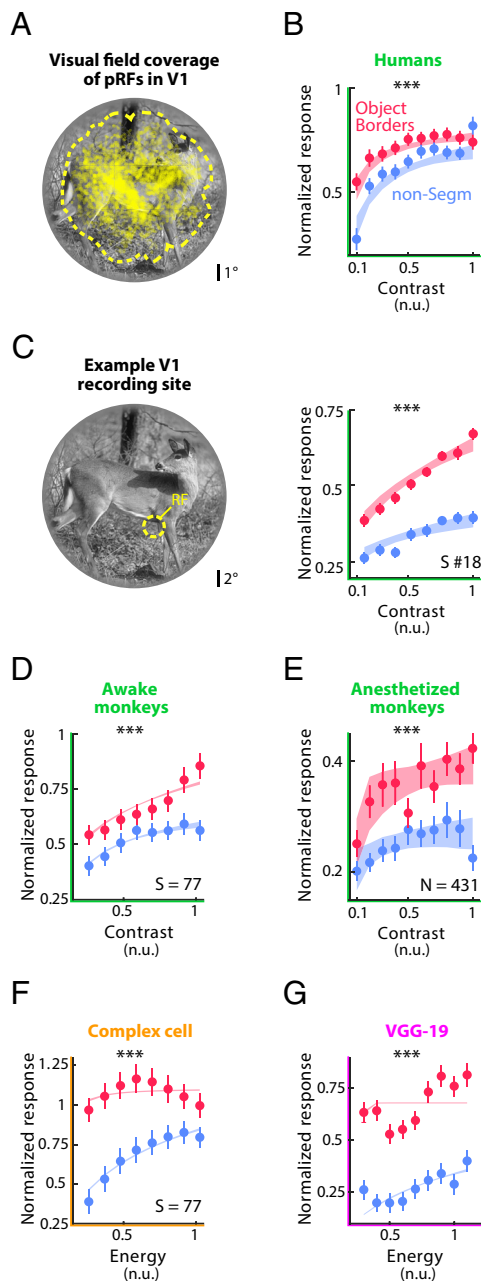
Fig. S1). To separate the influence of local contrast from that of object perception, we computed contrast response functions (CRFs; Fig. 1E), estimating the local contrast in the (population) receptive field (pRF) (18). Specifically, we compared activity in the same RMS contrast bin between stimuli with the (p)RF on object borders and nonsegmented contrast borders (*SI Appendix, Fig. S1 A and B*) (1, 19).

First, we collected data of four human participants who viewed 45 images from the BSD with 7 Tesla fMRI (*SI Appendix, Fig. S2*) (20). We mapped the entire visual field (Fig. 2A and *SI Appendix, Fig. S3*) and examined neuronal activity in multiple visual areas including V1, V2, V3, hV4, the lateral occipital visual field maps 1/2 (LO-1/2), and V3-a/b, comparing the activity elicited by image patches with and without object borders falling into the pRFs, at each cortical location (*SI Appendix, Fig. S3*) (21).

Across the contrast levels, the neuronal responses elicited by object borders were stronger than those elicited by nonsegmented contrast borders in V1 (Fig. 2B;  $P < 0.001$ , bootstrap test), and V2 and V3 (*SI Appendix, Fig. S3E*;  $P < 0.001$ , bootstrap test). We will refer to the extra activity elicited by object borders as "border modulation (BoM)." Furthermore, these results also held for the individual participants (*SI Appendix, Fig. S4*). The effects of perceptually relevant object borders were absent from the fMRI activity measured in areas V3ab, hV4, and LO-1 and 2 (all  $P_s > 0.05$ , bootstrap test; *SI Appendix, Fig. S3E*), which might be caused by the large pRF sizes in those areas, making the effect harder to detect (*Methods*). We conclude that the representation of object borders in natural images is enhanced in the early visual cortex of humans.

Second, we investigated how perceptually relevant borders influence multiunit spiking activity (MUA) in two awake macaque monkeys. We recorded the response elicited by four BSD stimuli, using chronically implanted electrode arrays in V1, while the monkey fixated at the center of the screen. The RFs of the V1 neurons were confined to a limited region of the visual field, but we recorded the neural responses to more than 500 different locations by changing the image position in each trial, effectively scanning the images with the RFs of 77 MUA recording sites (number of sites,  $S = 44$  in monkey B;  $S = 33$  in monkey M). After they fixated the center of the screen, the monkeys made a sequence of eye movements across the image, following sudden shifts of the fixation point across the image with their gaze (*SI Appendix, Fig. S1*). However, we focused our main analysis on the first fixation, when the image appeared on the screen.

Fig. 2C shows the RF (*Left*) and CRF (*Right*) of an example V1 recording site ( $S \#18$  in monkey B). Borders that were marked by the human observers elicited more activity than image regions of the same contrast that did not belong to object borders (time window 25 to 75 ms,  $P < 0.001$ , bootstrap test). We replicated this effect across the population of V1 recording sites. Object borders elicited stronger responses than nonborder image regions with the same local contrast in V1 (Fig. 2D;  $P < 0.001$ , bootstrap test). BoM was present at most recording sites in monkey B (64% of the sites) and monkey M (81% of the sites—time window: 25 to 75 ms; all  $P_s < 0.05$ , bootstrap test). We next analyzed the V1 activity elicited by the later fixations, during which a new part of the image appeared in the RF because the monkey shifted its gaze to a new position of the fixation point. These later fixations were preceded by saccades, causing a rapid movement of part of the image through the RF. The image was now familiar, and the monkey may have recognized and segmented the objects during the previous fixations. Accordingly, the BoM was also present when a saccade caused a border or nonborder image patch to fall in the



**Fig. 2.** Object borders enhance V1 responses in humans and both awake and anesthetized monkeys. (A) Visual field coverage of pRFs in V1: To examine the coverage of the images by V1, we plotted the pRF center positions (yellow dots) across all participants. The dotted line indicates the portion of the visual field covered by at least one pRF. (B) Normalized fMRI activity elicited by the object borders (red) and other image regions (blue) in human V1 (data averaged across 45 images and four participants). See *SI Appendix, Fig. S4* for the data of the individual participants. (C) Example V1 recording site. *Left*, RF of a representative recording site in an awake monkey (S #18 in monkey B). *Right*, normalized MUA responses elicited by the object borders (red) and other image regions as function of RMS contrast (blue). (D) Normalized MUA responses elicited by the object borders (red) and other image regions (blue) in area V1 as function of RMS contrast (data averaged across all recording sites in two awake monkeys). (E) Activity elicited by single neurons in V1 of anesthetized monkeys (data averaged across 53 images and two monkeys). (F and G) We binned the V1 activity in awake monkeys based on a complex cell model (F, orange) and VGG-19 model energy (G, purple). Note that the output of these models did not account for the difference in activity elicited by object borders (red) and other image regions (blue) (data averaged across four images and two awake monkeys). The equivalent result for the single units in anesthetized monkeys is shown in *SI Appendix, Fig. S1C*. Shaded regions denote 95% CI of the CRFs (determined by bootstrapping). Error bars indicate SEM across voxels (B) recording sites (D) or neurons (E). \*\*\* indicates  $P < 0.001$ , bootstrap test. S indicates the number of recording sites while N indicates the number of single neurons.

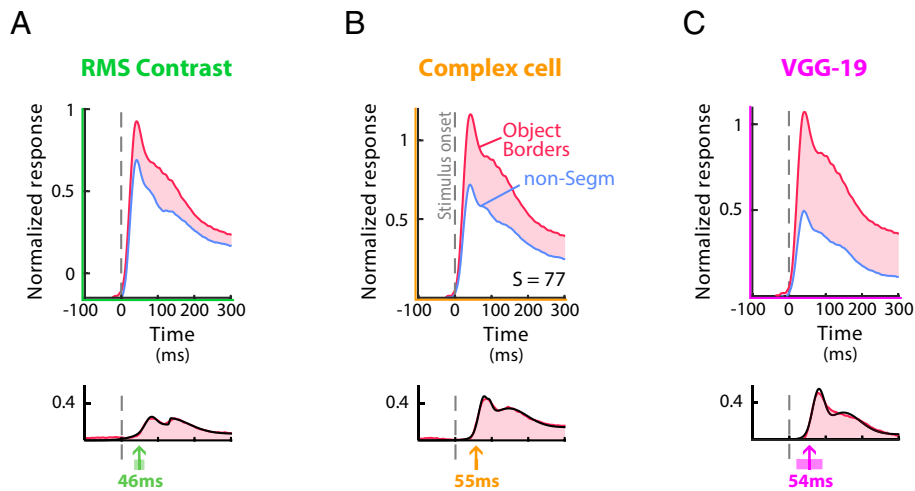
RF than if a new stimulus appeared on the screen (*SI Appendix, Fig. S1 E and F*;  $P < 0.001$ , bootstrap test).

Third, we investigated whether object borders influence single neuron responses in a published dataset with brief presentations (100 ms) of 53 BSD images, obtained in two anesthetized, paralyzed monkeys (22, 23). Anesthesia suppresses contextual modulations in V1 neurons (20), and our analysis permits a comparison to studies that examined V1 models using neuronal activity recorded during anesthesia (6, 23). Akin to the experiment in awake monkeys, object borders induced a fast (time window 25 to 75 ms) boost of V1 activity, which was independent of the local contrast in the RF (Fig. 2E,  $P < 0.001$ , bootstrap test). Thus, our results using human neuroimaging, multiunit activity in awake macaques, and single neurons in anesthetized macaques were convergent by revealing that object borders increase neuronal activity in the early visual cortex.

We wondered whether the BoM could be explained by current models of V1 tuning. We selected two models that predict the activity of V1 neurons: a model of complex cells and a model based on the convolutional VGG-19 network. The complex cell model is a classical description of the selectivity of V1 neurons to contrast, orientation, and spatial frequency (5) (see *Methods* for details; we also tested several advanced variants of this model in *SI Appendix, Supplementary Results*). The VGG-19 model is based on layer “conv3\_1” of VGG-19, which represents the state-of-the-art in predicting V1 responses to natural images (6). Strong BoM was also observed when we equated the orientation energy, i.e., the output of the complex cell model, between object borders and other image regions and we obtained the same result when we equated the output of the VGG-19 model in awake (Fig. 2 F and G; all  $P$ s  $< 0.001$ , bootstrap test) and anesthetized monkeys (*SI Appendix, Fig. S1 C*; all  $P$ s  $< 0.001$ , bootstrap test). Furthermore, we evaluated models that include cross-orientation suppression and surround suppression but observed that these models also do not account for the BoM (*SI Appendix, Supplementary Results*). Hence, our results, taken together, indicate that the BoM cannot be explained by classical or state-of-the-art models of V1 processing.

**The Latency of Object BoM in the Primary Visual Cortex.** To estimate the latency of the BoM, we examined the time course of V1 activity in awake monkeys (Fig. 3; for anesthetized monkeys, see *SI Appendix, Fig. S1D*), fitting a curve to the difference in activity elicited by the object borders and nonsegmented image borders, averaged across recording sites. We estimated latency as the time point at which the fitted function reached 33% of its maximum (*Methods*) (12, 13, 24). The BoM had a latency of  $46 \pm 15$  ms when sorting trials based on RMS contrast (mean  $\pm$  SD across bins) and the latency was similar when we binned using the complex cell model ( $54 \pm 6$  ms) or VGG-19 ( $55 \pm 34$  ms) (Fig. 3). Thus, object borders modulate the initial transient phase of the V1 response. This short latency of the BoM suggests that V1 neurons are tuned to features characterizing object borders in natural images, without the need of feedback from higher visual areas because these recurrent influences usually occur at a longer delay (7, 9, 10, 12).

**BoM in the Absence and Presence of Contextual Information.** The early timing of the V1 BoM indicates that it might be driven by the information inside the RF, unlike BoM in previous studies using artificial stimuli that held the stimulus in the RF constant but varied the context (7, 16) (Fig. 1F). We carried out additional experiments in awake monkeys to examine whether the early BoM reflects V1 tuning. In the first additional experiment, we focused on the influence from inside the RF by removing the information



**Fig. 3.** Perceptual borders enhance the early responses of V1 neurons. Normalized V1 response time courses in the center bin of the CRF equated for RMS contrast (A), the output of the complex cell model (B), or VGG-19 output (C). Lower panels show the difference in activity elicited by object borders and contrast/energy-matched nonborder regions. Colored arrows indicate the latency of BoM. For each model, object borders enhance MUA responses in V1 at a latency of ~50 ms. The equivalent results for single units in V1 are shown in *SI Appendix, Fig. S1D*.

outside the RF, and in the second experiment, we focused on the contextual influences by keeping the information inside the RF constant.

In the first additional experiment, we removed the context by copying circular image patches from the BSD that matched the V1 MUA RFs in size onto a gray background. We chose patches with the same RMS contrast that did or did not contain a segmented object border and centered them on the RFs of neurons at 50 recording sites in monkey B which was awake and fixating (Fig. 4A).

The isolated patches with object borders elicited a stronger V1 response than nonborder image regions with the same contrast ( $P < 0.001$ , Wilcoxon signed-rank test; Fig. 4B and C), with a BoM (i.e., the difference in response between object borders and nonobject image regions) up to more than 20%, in line with previous reports (12). We repeated this analysis also for the individual neurons of the two anesthetized monkeys (22). Again, responses elicited by image patches with object borders were stronger than those elicited by contrast-matched patches without segmented borders ( $P < 0.001$ , Wilcoxon signed-rank test; Fig. 4D). The median BoM of the single units was comparable to that of MUA in awake monkeys, but several neurons showed much stronger effects, up to more than 50% response enhancement. Hence, V1 neurons are tuned to features that uniquely characterize object borders, but these features are not captured by current V1 models. This form of V1 tuning explains part of the extra activity elicited by object borders.

The second additional experiment isolated possible contextual effects by placing the RF of 98 V1 recording sites (68 in monkey B and 30 in monkey M) on object borders and other locations in 12 natural images from the BSD, while keeping the image patch in the RF constant (Fig. 5A). Specifically, we copied an image patch with an object border and pasted it at a background location to create a condition in which the same image patch is not perceived as object border. An example image is shown in Fig. 5A, *Left* where we copied a part of the back of the elephant into the background.

On average, the object contours elicited a stronger V1 response than the same image patches presented at background locations ( $P < 0.001$ , Wilcoxon signed-rank test across recording sites, time window 0 to 300 ms; Fig. 5B). The latency of BoM in this experiment was 81 ms, i.e., it now occurred during the delayed phase

of the V1 response. The presence of a perceivable border increased the V1 response by 5%, on average, which is in line with the modulation observed using artificial textures (12). Thus, the context also influences the V1 response but at later time points, akin to experiments with artificial stimuli (8, 25, 26).

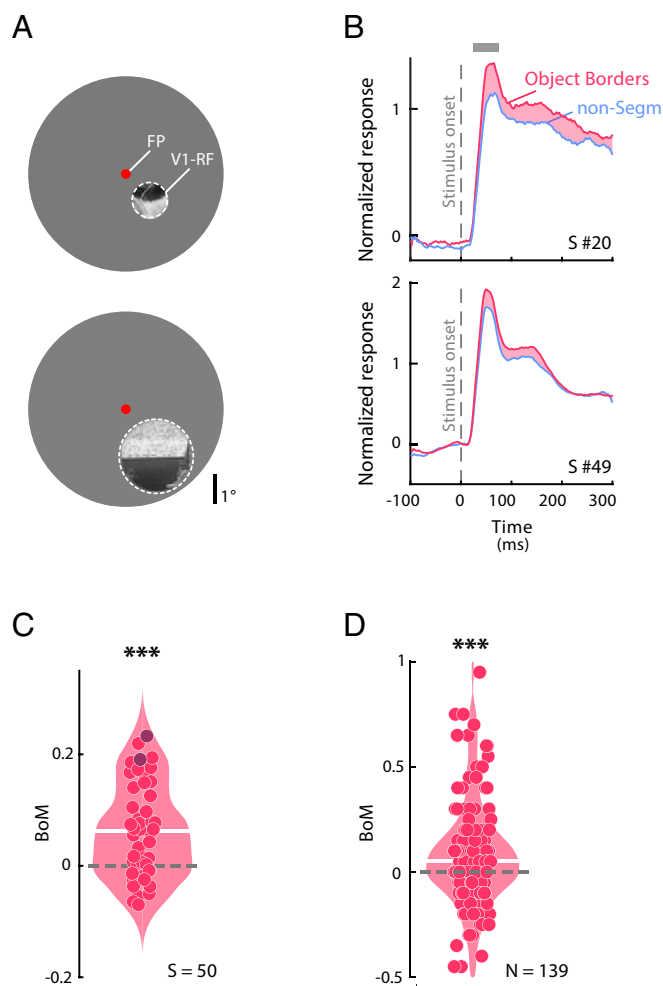
Finally, we aimed to remove the contextual effect by placing the image patches at identical locations of synthetic metamers of these images. The metamers had the same orientations, phases, spatial frequencies, auto- and cross-correlations, and marginal statistics, but the layout of objects was scrambled (27). In the example metamer of Fig. 5A, *Middle* and *Right*, the transitions between water, trees, and air were at the same locations, but the elephant was removed (other example metamers are shown in *SI Appendix, Fig. S5*).

BoM was absent for the metamers ( $P > 0.05$ , Wilcoxon signed-rank test; Fig. 5C). To investigate whether the level of BoM differed between the metamers and the original images, we performed a repeated-measures two-way ANOVA with object border and scrambling (two levels each) as factors. The main effects of object borders and scrambling were both significant (object border,  $F_{1,97} = 28.6$ ,  $P < 0.001$ ; scrambling,  $F_{1,97} = 5.42$ ,  $P = 0.022$ ). Importantly, the interaction was also significant at the population level ( $F_{1,97} = 6.74$ ,  $P = 0.011$ ) and for many of the individual recording sites (at  $P < 0.05$ ; 40% of the sites in monkey B and 73% in monkey M). Hence, if the RF stimulus is kept constant, contextual information can also enhance the V1 activity elicited by object borders, at a latency of ~80 ms.

These results, taken together, indicate that there are two processes that jointly explain the enhanced activity elicited by object borders. The tuning of V1 neurons enhances their representation from an early time point onward (~50 ms), and the scene context causes an additional activity increase at a longer latency (~80 ms).

## Discussion

We investigated how object borders in natural images influence the response of V1 neurons using human neuroimaging, recordings of multiunit activity in macaque monkeys, and single neurons in anesthetized monkeys. The results obtained with these methods were convergent by showing that perceptually relevant borders enhance neuronal activity in area V1. Indeed, the neuronal responses in V1 were similar between awake and anesthetized



**Fig. 4.** Fast, local BoM in V1 of awake and anesthetized monkeys. (A) Examples of isolated BSD image patches matching RFs of different V1 recording sites. (B) Time course of the V1 responses of two example recording sites. Object borders elicited stronger early activity than nonborder image patches. The gray horizontal line shows the 25 to 75 ms time window of interest. (C) Distribution of early (time window, 25 to 75 ms) BoM in awake monkeys elicited by image patches across recording sites. White bar, median BoM; \*\*\* $P < 0.001$ , Wilcoxon signed-rank test. S, number of recording sites; darker dots show the modulation of the two examples in (A). (D) Distribution of early BoM in anesthetized monkeys elicited by image patches across single neurons. N, number of single units.

monkeys, and between monkeys and humans (Fig. 2). We observed an influence of perceptually relevant borders on both the early V1 responses, which were driven by portions of the image that fell within the neurons' RFs, and we also observed a delayed influence of object borders, which was driven by image regions outside the RF.

**Early and Later Object Border Signals.** Unexpectedly, natural images elicited BoM during the initial V1 response at a latency of  $\sim 50$  ms. This is much earlier than in previous studies that used well-controlled, but artificial stimuli to keep the RF stimulus identical between relevant and nonrelevant contour conditions (7, 16). In these previous studies, the contextual effects on neuronal firing rates were attributed to feedback from higher cortical areas and/or lateral connections within V1, which can inform neurons about information outside the RF. The synaptic and propagation delays associated with these recurrent routes explain why BoM occurs a few tens of milliseconds after the initial V1 response (16, 25). Our results indicate that the early BoM signals evoked by natural images are not contextual but reflect the tuning of V1

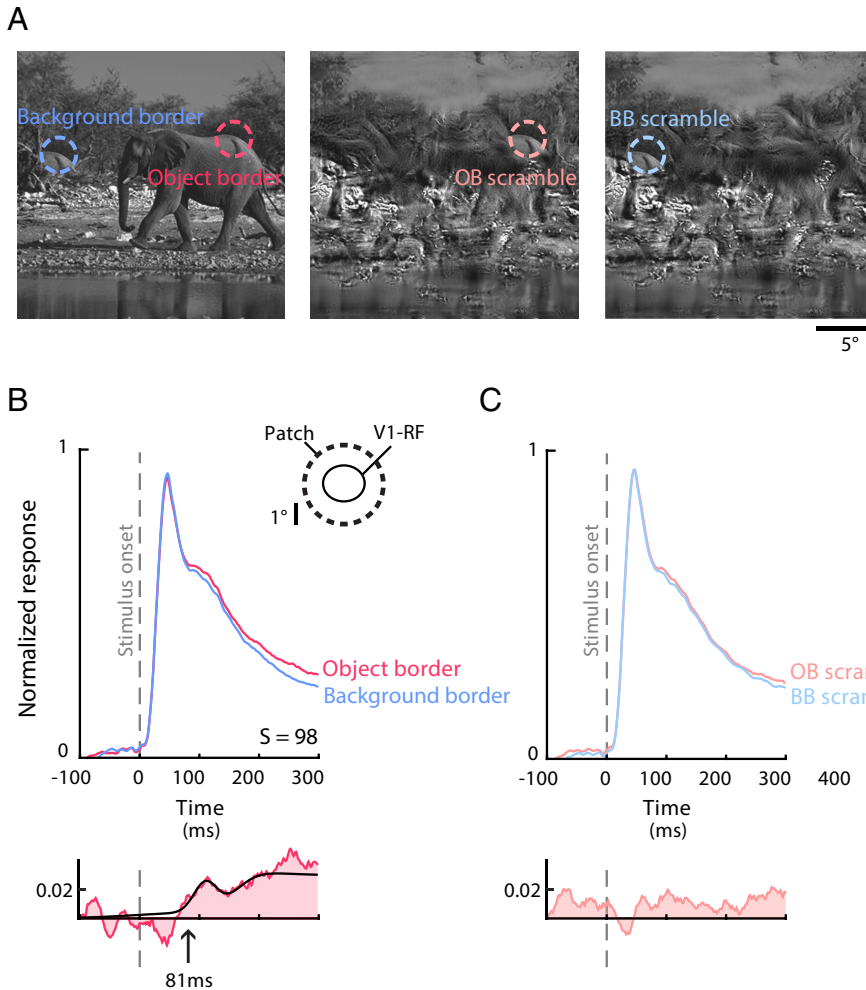
neurons. On average, the object borders elicit more activity than nonborder image patches with equated energy, independently from the specific model used to compute the energy. Apparently, these strong V1 responses to object borders are more complex than can be described by complex-cells Gabor filters (6, 28, 29) or divisive normalization (*SI Appendix, Supplementary Results*), i.e., "classical" models of V1 tuning (30) and were not predicted by existing models of V1 tuning, including state-of-the-art models based on AI (6, 29). We replicated the tuning to object borders by showing isolated image patches (Fig. 4) and in a dataset in anesthetized monkeys, supporting the view that these effects do not depend on feedback, because contextual feedback influences are diminished under anesthesia (20, 31). Future research is needed to understand how these influences are produced by the connectivity from the LGN to V1 and the local connectivity inside cortical columns in area V1 (32, 33).

In addition to their effect on the feedforward response, object borders also elicited a contextual influence on V1 activity. When we matched the image elements of object and nonobject contours in the RF of V1 neurons, the activity elicited by the object contours was still stronger than that elicited by other, nonobject contours (Fig. 5). BoM now occurred at a latency of 81 ms, which is 30 ms later than the feedforward response, which is in line with previous studies that used synthetic stimuli to keep the RF content constant while controlling contour saliency by the layout of image elements in the surround (7, 16). This additional delay suggests that contextual BoM depends on feedback from higher areas and/or horizontal connections within V1. It is of interest that these putative feedback signals increased the activity elicited by contours that are predicted by an object's overall shape. This result is not in accordance with popular "predictive coding" schemes (34), which suggest that feedback connections should suppress the activity of contours that are predicted by the shape of the object. Instead, we found that object borders increase the neuronal activity in the visual cortex, both during the early and later phases of the V1 response.

BoM is presumably related to border-ownership coding, which is expressed by many neurons in V2, V3, V4, and by some V1 neurons, and it also occurs for natural images (35–37). The activity of neurons with border-ownership signals depends on the side of the figural region relative to the border that falls in the RF. For example, if the border is vertical, some neurons prefer stimuli in which the border in their RF is owned by a figure to the left, whereas other neurons have the opposite preference. Hence, border-ownership neurons can link the shape of the border to the surface properties of the object's interior, and they may therefore play an important role in object recognition and segmentation. In many situations, the local shape of a border falling in a RF can provide information about the side of the figure (38). In these situations, neurons express border-ownership early, during the feedforward response. However, if the RF stimulus is kept constant, border-ownership coding occurs after an additional delay (37). Our design did not allow us to measure border-ownership tuning, but it is probable that some of the extra activity elicited by the object borders observed by us is related to the neurons' preferred figure-side and, hence, that the two effects are intimately related.

## Conclusion

We conclude that the object borders in natural images increase the response of V1 neurons. The extra neuronal activity occurs during the early response phase if the local image elements in the RF are indicative of an object border and later in time if it depends



**Fig. 5.** Contextual BoM of the delayed V1 response. (A) To examine the role of contextual information around the V1 RF, we modified natural images ensuring that the same features were present the RFs. We either copied an image patch with an object border to a background location (Left) or removed the object from the scene by creating *metamers* (Middle and Right). The RF stimulus was identical in all conditions. (B) Average V1 response elicited by image regions that demarcated object borders (red; CI were too small to visualize) or were part of the background (blue). Lower panel, response difference. BoM in this experiment had a latency of 81 ms. The relative size of the patch, with respect to the median V1-RF, is shown in the Top Inset: All V1-RFs were well inside the patch. (C) The responses elicited by the metamers did not reveal significant differences.

on contextual information outside the RF. The selectivity for object borders represents an unexpected dimension of V1 tuning that goes beyond oriented filter models and state-of-the-art deep networks. Future studies may now further explore the features that characterize object borders detected in V1, thereby aligning the mechanistic functions of this brain region with its critical role in object perception.

## Methods

### RMS Contrast and Models of V1 Tuning.

**RMS contrast.** To derive the RMS contrast, we computed the contrast of every natural image within each site/neuron's RF or voxel's pRF. The (p)RF was modeled as a circular symmetric Gaussian function, described by parameters for position  $(x_c, y_c)$  and size  $(\sigma)$ , giving rise to a Gaussian weighting function  $w_i$ :

$$w_i = \exp\left(-\frac{(x_i - x_c)^2 + (y_i - y_c)^2}{2\sigma^2}\right), \quad [1]$$

where  $x_c$  and  $y_c$  define the location of the center of the (p)RF in the visual field,  $\sigma$  determines the size of the (p)RF, and  $x_i$  and  $y_i$  define the location of the  $i$ -th pixel of the image. We computed each site/neuron/voxel's contrast value to each natural image by calculating the RMS contrast (39, 40) of the part of the image inside their (p)RF. RMS contrast was defined as the SD of the luminance of the

pixels relative to the mean. The RMS-contrast was weighted by the (p)RF Gaussian function to obtain the local contrast-energy value:

$$\text{RMS contrast energy} = \sqrt{\frac{1}{\sum_{i=1}^N w_i} \sum_{i=1}^N w_i \frac{(L_i - L)^2}{L^2}}, \quad [2]$$

where  $N$  is the number of pixels in the stimulus window,  $L$  is the mean luminance of the pixels inside the spatial window,  $L_i$  is the luminance of the  $i$ -th pixel (2), and  $w_i$  comes from Eq. 1.

**Complex cell model.** The magnitude of the complex-cell response was computed as

$$\text{Complex cell energy} = \sqrt{E_{x,y,F,\theta,\Phi=0}^2 + E_{x,y,F,\theta,\Phi=90^\circ}^2} \quad [3]$$

where  $E$  is the dot product of the stimulus (grayscale) pixel luminance with a 2D Gabor with coordinates  $x$  and  $y$ ,  $F$  is spatial frequency,  $\theta$  orientation, and  $\Phi$  phase. For neurons in anesthetized monkeys,  $F$  and  $\theta$  were computed using a separate set of stimuli containing Gabor patches with varying spatial frequency, orientation, and phase. For recording sites in awake monkeys and fMRI voxels,  $F$  and  $\theta$  were free parameters of the model and were optimized for every recording site before we calculated the explained variance. We used Matlab's *lsqcurvefit* function to compute the fits.

**VGG-19.** We extracted the activity of units of VGG-19's layer conv3\_1, which is state of the art in predicting V1 responses to natural images (6, 29), and followed the

approach of ref. 6 with two modifications. We used a two-step mapping (41, 42), described by following the equation:

$$r_{VGG-19energy} = f_{VGG-19}(input) * W_s * W_d, \quad [4]$$

where  $r_{VGG-19energy}$  is the predicted response of a V1 recording site/neuron,  $f_{VGG-19}$  ( $input$ ) is the output of VGG-19's conv3\_1 to our stimulus set (i.e.,  $input$ ), and  $W_s$  and  $W_d$  are two sets of weights defining spatial and feature selectivity, respectively. The spatial mask ( $W_s$ , set as the 2D Gaussian RF estimate) approximates the RF and a weighted sum of the nodes in the ANN ( $W_d$ ) approximates the feature selectivity of the recorded sites/neurons (41).  $W_s$  and  $W_d$  were determined separately for every recording site/neuron, while training occurred at the same time for all of them. We trained the model to optimize  $W_d$  to predict V1 responses to the training set (i.e.,  $input$  in Eq. 4). We cross-validated the model for each site/neuron and used the final model to extract VGG-19 energy for each trial. For recording sites in the awake monkeys, we cross-validated using random 50% splits of the trials. For neurons in the anesthetized monkeys, we used 217 images for training, and 27 images for testing. For the anesthetized monkeys, the training images were fixed and the test sets differed across neurons because the test images depended on the presence of object borders in the RF. To avoid overfitting, we fitted models (i.e., optimized  $W_d$ ) using images within the 25th-75th percentiles of object border relevance (see below in the section on quantification of object borders) and used the other images during testing. We tested the object BoM using VGG-19 energy with the 27 held-out stimuli, only including neurons for which VGG-19 energy predicted the neural responses on the test images ( $N = 26$ , Pearson's correlation  $r > 0.323$ ,  $P < 0.05$  uncorrected,  $t$  test).

**CRFs.** For RMS contrast (and for the models of tuning), we computed the CRF of the MUA recording sites and single neurons in V1, and voxels of areas V1, V2, V3, hV4, LO-1/2, and V3-a/b (see below) by measuring their responses as a function of local contrast (or the model energy) inside their (p)RF. We chose energy bins such that every bin spanned 10% of the contrast (or energy) distribution and fitted the following equation (modified from ref. 43):

$$R(C) = a \frac{C^q}{C^q + Q^q}, \quad [5]$$

where  $R$  is the neural response,  $C$  is the contrast (or model energy) inside the pRF,  $Q$  represents the model energy value where  $R$  is at half of its maximum response, and  $q$  determines the slope ( $a$ ,  $Q$ , and  $q$  are free parameters). The number of bins did not influence the outcome.

**Quantification of object borders.** The BSD images have been annotated by 5 to 9 human observers who drew lines to segment object borders that are relevant for the scene's interpretation (1, 19). We used these measurements to define the perceptually relevant borders of the scene. Every pixel  $i$  of the manually labeled images has a value for the confidence of observers that the pixel belongs to an object border,  $S_i$ , between 0 (not labeled by any observer) and 1 (labeled by all observers). The border relevance in the pRF is calculated as a weighted sum across pixels that fall in it:

$$\text{Perceptual border relevance} = \frac{\sum_{i=1}^N w_i \cdot S_i}{\sum_{i=1}^N w_i}. \quad [6]$$

Here,  $w_i$  are the weights of the (p)RF estimate (Eq. 1), and  $N$  is the total number of pixels in the (p)RF. For every (p)RF, we computed the distribution of border relevance across images, and we included the lowest quartile of the distribution as "nonborder patches" and the highest quartile as "object borders." We then computed the CRFs within these classes (SI Appendix, Fig. S1). These thresholds were not critical because we replicated the results when we included all responses, i.e., with a median split (not shown).

**Statistics.** We used a bootstrapping procedure to determine the significance of differences in CRFs between conditions. We sampled the images with replacement 1,000 times, fit the CRF for the two simulated conditions, and computed the mean difference in the area under the curve of the two conditions. We derived the  $P$ -value from this distribution of differences.

#### Electrophysiological Experiments in Awake Monkeys.

**Training of the monkeys.** All procedures complied with the NIH Guide for Care and Use of Laboratory Animals and were approved by the institutional animal care and use committee of the Royal Netherlands Academy of Arts and

Sciences. Two macaque monkeys (males, 7 and 13 y old) participated in the electrophysiological experiments. They were socially housed in stable pairs in a specialized primate facility with natural daylight, controlled humidity, and temperature. The home cage was a large floor-to-ceiling cage which allowed natural climbing and swinging behavior. The cage had a solid floor, covered with sawdust, and was enriched with toys and foraging items. Their diet consisted of monkey chow supplemented with fresh fruit. Their access to fluid was controlled, according to a carefully designed regime for fluid uptake. During weekdays the animals received water or diluted fruit juice in the experimental set-up upon correctly performed trials. We ensured that the animals drank sufficient fluid in the set-up and supplemented extra fluid after the recording session if they did not drink enough. On days of the weekend, they received at least 700 mL water in the home cage in a drinking bottle. The animals were regularly checked by veterinary staff and animal caretakers and their weight and general appearance were recorded daily in an electronic logbook during fluid-control periods.

**Surgical details.** We implanted both monkeys with a titanium headpost (Crist instruments) under aseptic conditions and general anesthesia as reported previously (44). The monkeys were trained to direct their gaze to a 0.5° diameter fixation dot and hold their eyes within a fixation window (1.1° diameter). They then underwent a second operation to implant 5 × 5 arrays of microelectrodes (Utah-probes, Blackrock Microsystems) over opercular V1. The interelectrode spacing of the arrays was 400 μm. We obtained good signals from 4 V1 arrays in each monkey (13).

**Electrophysiology in awake monkeys.** We recorded neuronal activity of 192 recording sites in V1 (96 in Monkey M and 96 in Monkey B). We recorded the envelope of multiunit activity by digitizing the signal referenced to a subdural electrode at 24.4 kHz. The signal was band-pass filtered (2nd order Butterworth filter, 500 Hz to 5 KHz) to isolate high-frequency (spiking) activity. This signal was rectified (negative becomes positive) and low-pass filtered (corner frequency = 200 Hz) to produce the envelope of the high-frequency activity, which we refer to as MUA (45). The MUA signal reflects the population spiking of neurons within 100 to 150 μm of the electrode and the population responses are very similar to those obtained by pooling across single units (45–48).

**Selection of recording sites and inclusion of data.** To normalize MUA, we first subtracted the mean activity in the pretrial period in which the animal was fixating (200 to 0 ms relative to stimulus onset) and divided by the maximum smoothed (26 ms Gaussian kernel) peak response (0 to 150 ms after stimulus onset). Neuronal activity was normalized to the peak response elicited by stimulus onset. The data are therefore in normalized units, where, e.g., a value of 0.1 indicates 10% of the maximal MUA onset response. We only included recording sites on days with a sufficient signal-to-noise ratio ( $SNR_{DAY}$ ).  $SNR_{DAY}$  was estimated by dividing the maximum of the initial peak response by the SD of the baseline activity across trials. When the  $SNR_{DAY}$  of a recording site was smaller than 2 on a particular day, we removed that session from the analysis of that recording site. Approximately, 80% of recording sites in monkey B and 77% in monkey M were included. To test for statistical differences between conditions and to compute the CRFs, MUA was averaged in a time window 25 to 75 ms after stimulus onset.

**Stimulus presentation.** In the experiments with awake monkeys, stimuli were presented on a CRT monitor at a refresh rate of 60 Hz and resolution of 1,024 × 768 pixels viewed from a distance of 46 cm. The monitor had a width of 40 cm, yielding a field-of-view of 41.6 × 31.2°. All stimuli were generated in Matlab using the COGENT graphics toolbox (developed by John Romaya at the LON at the Wellcome Department of Imaging Neuroscience). The eye position was recorded using a digital camera (Thomas recordings, 250 Hz frame rate).

**Receptive field mapping.** We mapped the RF of each MUA recording site in V1 of the awake monkeys using a drifting luminance-defined bar that moved in one of four directions. The response to each direction was fitted with a Gaussian function. The borders of the RF were then calculated as described previously (45). The signal-to-noise ratio ( $SNR_{RF}$ ) of the response was taken as the peak of the Gaussian divided by the SD of the pretrial baseline response. We only included recording sites in the analyses with a reliable visual response (i.e., the responses to all four bar directions had an  $SNR_{RF}$  of at least 1). The median V1 RF size, taken as the square root of the area, was 1.8° (range 0.4° to 8.2°) and the median eccentricity of the RFs was 2.4° (range 0.6° to 12.9°) (49).

**Analysis of latency.** To compute the latency of the difference between the response to object borders and nonsegmented borders, a function was fitted to the time course of the difference response (12, 13, 24). The function was derived from the assumptions that the onset of the response difference has a Gaussian distribution and that a fraction of the response dissipates exponentially which yields the following equation:

$$f(t) = d \cdot \exp(\mu\alpha + 0.5\sigma^2\alpha^2 - \alpha t) \cdot (G(t, \mu + \sigma^2\alpha, \sigma) + c \cdot G(t, \mu, \sigma)), \quad [7]$$

where  $G(t, \mu, \sigma)$  is a cumulative Gaussian density with mean  $\mu$  and  $SD\sigma$ ,  $\alpha^{-1}$  is the time constant of the dissipation, and  $c$  and  $d$  represent the contribution the nondissipating and dissipating components, respectively. The latency was defined as the point at which the fitted function reached 33% of its maximum.

**Natural images presented in the awake electrophysiological experiments.** Four BSD images were used in the electrophysiological experiments (11.6° radius visual angle; *SI Appendix, Fig. S2*). At the start of the trial, the screen was gray (26.8 cd m<sup>-2</sup>) with a red fixation point with a position that was randomly selected from uniformly spaced grid (with ~500 positions) covering the circular aperture of the image. The image appeared once the monkey had maintained fixation for 300 ms (fixation 1). After an additional 400 ms, the first fixation point disappeared and the next fixation points appeared, and these shifts of the fixation point occurred five times per trial. Reward was delivered after every correct fixation, with an extra amount at the end of the trial. Here, we mainly focused on the data during the first fixation to facilitate the comparison with the data from anesthetized monkeys and human fMRI. We collected a total of 11,783 correct trials for monkey M and 13,373 for monkey B.

**Statistics.** We compared differences between the CRFs or model output between object borders and nonborder patches using a bootstrapping procedure (1,000 iterations), as described above.

**Electrophysiological Experiments in Anesthetized Monkeys.** We analyzed neural responses from a public dataset (22, 23) that was collected in three anesthetized and paralyzed macaque monkeys. We included the responses to 431 spike-sorted neurons with a reliable RF estimation ( $R^2 > 0.5$ ) from monkeys 1 and 2. We excluded the data of monkey 3, for which a smaller version of the stimuli was used. Details on surgery, spike sorting, and stimulus presentation are described in full in the original publication (23). In brief, the authors used 956 stimuli, which were shown for 100 ms each (20 repetitions). The stimulus set included 270 full natural images (diameter, 6.7 dva), 270 patches of the images (circular crops with a diameter of 1.04 dva), and 416 circular Gabor patches (diameter, 1.04 dva), which were used to estimate tuning to orientation and spatial frequency. Of the 270 full natural images, we selected the 53 images that came from the BSD and had been shown in their original orientation. We summed their responses in the 25 to 75 ms time window and then averaged across repetitions. For the isolated patch analysis, we used the patches that came from the same 53 stimuli. The Gabor patches and the patches of the natural images fell into the RF of 139 of the 431 neurons. Hence, for these cells, we could compute complex cell energy and we could also include them in the isolated patch experiment.

#### fMRI Experiment with Human Participants.

**Participants.** Four participants (all male; ages 29 to 41 y) participated in the fMRI experiment. All participants had normal or corrected-to-normal visual acuity. We obtained informed written consent of the participants and the protocol was approved by the Human Ethics Committee of University Medical Center Utrecht.

**Stimulus presentation.** The visual stimuli were generated in Matlab (Mathworks Inc.) using the PsychToolbox (39, 50) on a Macintosh MacBook Pro. The stimuli were backprojected on a display inside the MRI bore. The subject viewed the display through mirrors inside the scanner. The size of the display was 15.0 × 7.9 cm with a resolution of 1,024 × 538 pixels. The total distance from the subject's eyes to the display was 41 cm. The stimuli were constrained to a circular area (radius, 5.5°) with the size of the vertical dimension of the screen. The area outside this circle was maintained at a constant mean luminance.

**Functional imaging and processing.** The MRI data were acquired with a Philips 7T scanner using a 32-channel head coil (18). We scanned the participants with a 2d-echo-planar-imaging sequence with 25 slices oriented perpendicular to the calcarine sulcus with no gap. The following parameters were used: repetition time

(TR) = 1,500 ms, echo time (TE) = 25 ms, and a flip angle of 80°. The functional resolution was 2 × 2 × 2 mm and the field of view (FOV) was 190 × 190 × 50 mm. We used foam padding to minimize head movement. The functional images were corrected for head movement between and within the scans. For computation of the head movement between scans, the first functional volumes for each scan were aligned. Within-scan motion correction was then computed by aligning the frames of a scan to the first frame. The duration of the pRF mapping scans was 372 s (248 time frames), of which the first 12 s (8 time frames) were discarded to eliminate start-up magnetization transients. During the three sessions, we acquired 6 to 8 pRF mapping scans in total per subject. To increase the signal-to-noise ratio, we averaged across the repeated scans. During the three sessions in which we presented the natural images, we acquired 6 to 7 scans for each of the three stimulus sets. The duration of the scans with the natural images was 432 s (288 time frames). The first 12 s (8 time frames) were discarded to eliminate start-up magnetization transients. The images were presented in a block design. Each image was presented during a 9-s block. Within this block, the same image was shown 18 times for a duration of 300 ms followed by 200 ms mean luminance. The full-field stimuli were presented with three alternating different high-contrast patterns, to obtain a full high-contrast response that is not based upon one specific high-contrast pattern (*SI Appendix, Fig. S3B*). Specifically, the phase of the full-field pattern was randomized on different presentations in order to obtain a response that is not influenced by one specific dartboard pattern. The block in which the stimulus was presented was followed by a 12-s mean luminance presentation. Four longer blank periods of 33 s were also included during the scan.

**Anatomical imaging and processing.** The T1-weighted MRI images were acquired in a separate session using an 8-channel SENSE head coil. The following parameters were used: TR/TE/flip angle = 9.88/4.59/8. The scans were acquired at a resolution of 0.79 × 0.80 × 0.80 mm and were resampled to a resolution of 1mm3 isotropic. The functional MRI scans were aligned with the anatomical MRI using an automatic alignment technique (51). From the anatomical MRI, white matter was automatically segmented using the FMRIB's Software Library (FSL) (52). After the automatic segmentation, it was hand-edited to minimize segmentation errors (53). The gray matter was grown from the white matter to form a 4 mm layer surrounding the white matter. A smoothed 3D cortical surface can be rendered by reconstruction of the cortical surface at the border of the white and gray matter (54).

**pRF mapping stimulus.** We used bar apertures filled with natural images (18, 21) (*SI Appendix, Fig. S3A*) to train the pRF model of participants in the fMRI experiment. The width of the bar subtended 1/4th of the stimulus radius (1.375°). Four bar orientations (0°, 45°, 90°, and 135°) and two different step directions for each bar were used, giving a total of 8 bar directions within a given scan. The bar stepped across the stimulus aperture in 20 steps (with a distance of 0.55° and a duration of 1.5 s per bar position) so that each pass took 30 s. A period of 30 s mean luminance (0% contrast) was presented after every pass. In total, there were 4 blocks of mean luminance during each scan, presented at evenly spaced intervals. The participants performed a fixation dot task to make sure they fixated at the center of the display. A small fixation dot (0.11° radius) was presented in the middle of the stimulus. The fixation dot changed its color from red to green at random time intervals and subjects were instructed to respond to color changes using a button press.

**Natural images.** The natural images came from the BSD (1, 19). The original resolution of the images was 321 × 481 pixels (both landscape and portrait). In the fMRI experiments (18), we selected a square region of 321 × 321 pixels from the images and upsampled it to a resolution of 516 × 516 pixels, which corresponds to a stimulus of 11 × 11° diameter of visual angle. The images were masked by a circle with a raised cosine faded edge (width of 0.9°), and the areas outside this circle were set to the mean luminance. The images were gamma-linearized and the mean contrast was set to 50%. We used 3 image sets in different scanning runs, each containing 15 different natural images (45 in total) and one full-field binarized bandpass-filtered noise stimulus. *SI Appendix, Fig. S2* shows the image set. A fixation dot was presented at the center of the stimulus. We used the same fixation dot task as for the pRF mapping runs.

**pRF model-based analysis.** The pRF model was estimated for every cortical location from the measured fMRI signal that was elicited by the pRF mapping bar stimuli (*SI Appendix, Fig. S3A*) (18, 21). In short, the method estimates the



pRF by combining the measured fMRI time series with the position time course of the visual stimulus. A prediction of the time series is made by calculating the overlap of the pRF and the stimulus energy (RMS contrast, see below) convolved with the hemodynamic response function (HRF). We estimated the parameters of the HRF that best describes the data of the whole acquired fMRI volume (55). The optimal parameters of the pRF model are chosen by minimizing the residual sum of squares between the predicted and the measured time series. We used the conventional pRF model, which consists of a circular symmetric Gaussian. This model has four parameters: position ( $x, y$ ), size ( $\sigma$ ), and amplitude ( $\beta$ ). For further technical and implementation details, see ref. 21.

This approach works best in areas with small (p)RFs, such as V1. In higher areas, the probability that pRFs contain a border is higher, because they are larger. Some of the pRFs in higher visual areas contain, for example, a short piece of object border, whereas the remainder was without any border, weakening the contrast. In other words, the distribution of object borders has a lower variance in higher areas, and this reduces the contrast between pRFs with and without object borders.

**Regions of interest.** We used the pRF method to estimate position parameters  $x$  and  $y$  of the pRF of every voxel. From these values, we derived the polar angle [ $\text{atan}(y_0/x_0)$ ] and eccentricity [ $\sqrt{x_0^2 + y_0^2}$ ] values. We drew the borders between visual field maps on the basis of polar angle and eccentricity maps on the inflated cortical surface (56). We defined visual areas V1, V2, V3, hV4, LO-1/2, and V3-a/b as our regions of interest (ROIs) (57–60).

**Analysis of fMRI responses to the natural images.** We measured fMRI responses to 45 natural images (SI Appendix, Fig. S2) and 3 full-field high-contrast stimuli (100% contrast; SI Appendix, Fig. S3B) (18). We first determined the voxel response amplitudes in %BOLD signal change elicited by each of these images. The voxel responses were calculated using a general linear model (GLM) (61, 62). To correct for differences in response amplitudes between voxels, we normalized the responses to the voxel's response to the full-field (100% contrast) stimulus.

To determine the CRF in the fMRI experiment, we only used the voxels with an overall significant response ( $t$ -values  $> 4.0$ ), a pRF eccentricity between 0.5 and 4° and for which the pRF model explained more than 40% of the variance. Based on previous work, for every area we used a threshold for the pRF sizes (21, 55, 63, 64). In V1, we included pRFs with a value of  $\sigma$  (which determines pRF size) between 0.25° and 0.8°, for V2 between 0.25° and 1.1°, for V3 between 0.25° and 1.75°, for hV4 between 0.45° and 3°, for V3-a/b between 0.45° and 3.75° and for LO-1/2 between 0.9° and 5°.

**Isolated Patch Experiment.** To test whether isolated image patches from the BSD that either contained object borders or not, elicited a different level of V1 activity, we carried out an additional experiment in monkey B (50 recording sites, Fig. 4 A–C) and a further analysis on data from two anesthetized monkeys (139 neurons, Fig. 4D).

In the awake monkey, we chose three V1 recording arrays and centered 100 patches of the image from the BSD that contained object borders and 100 patches that did not on the RFs. These patches were selected so that the RMS contrast was the same ( $70 \pm 1\%$ ) and the size matched the median RF of the recording sites of the array (0.9° to 2.0°). The patches were presented on a gray background ( $26.8 \text{ cd m}^{-2}$ ) while the monkey maintained gaze on a red fixation point for 300 ms. We repeated each stimulus five times and collected a total of 3,000 trials (1,000 trials per array).

In the published dataset from anesthetized monkeys, we included neurons for which isolated patches (1.04 dva) were well centered inside the RF ( $N = 139$ ). The presentation duration was 100 ms. For each neuron, we selected pairs of patches with the same RMS contrast ( $\pm 1\%$ ) with and without object borders in the RF and computed the average response. We then averaged responses across all neurons that had been included.

In both datasets, we tested the significance of the difference in the activity elicited by isolated object and nonsegmented border patches during the peak of the response (25 to 75 ms) with a Wilcoxon signed rank test across recording sites and neurons.

**Contextual BoM Experiment.** To examine differences in activity elicited by object borders and nonborder patches in V1 of awake monkeys when the stimulus in the RF was held constant (Fig. 5), we selected twelve images from the BSD, which were cropped and upsampled to  $512 \times 512$  pixels ( $23.2^\circ \times 23.2^\circ$ ). We ensured that the portion of the image covered by the RF of each recording site and its surround were exactly the same across conditions (same size and content, Fig. 5), so that border salience only depended on information outside the neurons' RF. We used a  $2 \times 2$  design. The first factor was whether the image element in the RF fell on an object border (Fig. 5A). The second factor was whether we presented the original image or a scrambled version (also known as *metamer*). To this aim, we created three further stimuli from each image. First, we copied a circular patch (80 pixels in diameter,  $3.7^\circ$ ) from an object contour location onto a nonobject contour location using Adobe Photoshop (blue circle in Fig. 5A, see SI Appendix, Fig. S5 for other example images). The border of this circular patch was smoothed to blend it in at the new location. We created two metamers using the algorithm of ref. 65, with Matlab code provided by the authors (<https://github.com/freeman-lab/metamers>). The two metamers were constructed so that either the object- or nonobject patch was kept intact, with a smooth transition to the surround.

Trials started with a red fixation point and the stimulus appeared after 300 ms of fixation. The monkeys maintained fixation for an additional 400 ms after stimulus onset. We ensured that the RFs of V1 recording sites were centered on the image patch, which was identical in the four conditions. The order of the conditions was randomized across trials and aborted trials (when the monkeys broke fixation) were repeated at the end. We collected a total of 8,094 trials in monkey M and 9,111 in monkey B.

We tested the significance of the BoM in a window from 0 to 300 ms after stimulus onset (subtracting spontaneous activity,  $-100$  to 0 ms) with a Wilcoxon signed rank test across recording sites. We also used a repeated-measures two-way ANOVA across recording sites, with object/nonobject contour and scrambled/not scrambled as factors.

**Data, Materials, and Software Availability.** Multiunit activity data and code have been deposited in OSF (DOI: [10.17605/OSF.IO/QPC2D](https://doi.org/10.17605/OSF.IO/QPC2D)) (66). Previously published data were used for this work (22).

**ACKNOWLEDGMENTS.** We thank Kor Brandsma and Anneke Ditewig for bio-technical support. The work was supported by the European Union's Horizon 2020 and FP7 Research and Innovation Program (Framework Partnership Agreement No. 650003 [Human Brain Project Framework Partnership Agreement], European Research Council advanced grant 101052963 "NUMEROUS" and grant agreement 899287 "NeuraViper"), "DBI2," a Gravitation program of the Dutch Ministry of Science, Education and Culture) and the Netherlands Organization for Scientific Research (NWO) Crossover Program 17619 "INTENSE" to P.R.R., the European Union's Erasmus+ program (2018-1-IT02-KA103-047276/10), the NWO Open-Competition Domain Science-XS (OCENW.XS22.2.097) and the NWO Veni (VI.Veni.222.217) to P.P. and the NWO Vidi (452.08.008) and Vici (016.vici.185.050) to S.O.D.

Author affiliations: <sup>a</sup>Department of Vision and Cognition, Netherlands Institute for Neuroscience (KNAW), Amsterdam 1105 BA, Netherlands; <sup>b</sup>Momilab Research Unit, Institutions, Markets, Technologies School for Advanced Studies Lucca, Lucca 55100, Italy; <sup>c</sup>Department of Computational Cognitive Neuroscience and Neuroimaging, Netherlands Institute for Neuroscience (Koninklijke Nederlandse Akademie van Wetenschappen), Amsterdam 1105 BA, Netherlands; <sup>d</sup>Spinoza Centre for Neuroimaging, Amsterdam 1105 BK, Netherlands; <sup>e</sup>Department of Integrative Neurophysiology, Vrije Universiteit Amsterdam 1081 HV, Netherlands; <sup>f</sup>Department of Neurosurgery, Academic Medical Centre, Amsterdam 1100 DD, Netherlands; <sup>g</sup>Laboratory of Visual Brain Therapy, Institut National de la Santé et de la Recherche Médicale, Centre National de la Recherche Scientifique, Institut de la Vision, Sorbonne Université, Paris F-75012, France; <sup>h</sup>Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam, Amsterdam 1181 BT, Netherlands; and <sup>i</sup>Department of Experimental Psychology, Helmholtz Institute, Utrecht University, Utrecht 3584 CS, Netherlands

Author contributions: P.P., W.Z., M.W.S., P.R.R., and S.O.D. designed research; P.P., W.Z., R.R.M.T., and A.G. performed research; P.P., W.Z., R.R.M.T., and M.W.S. analyzed data; and P.P., W.Z., M.W.S., P.R.R., and S.O.D. wrote the paper.

1. D. Martin, C. Fowlkes, D. Tal, J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics" in *Proceedings of the IEEE International Conference on Computer Vision. ICCV 2001* (IEEE, 2001), pp. 416–423.
2. W. Zuidersbaan, J. van Leeuwen, S. O. Dumoulin, Change blindness is influenced by both contrast energy and subjective importance within local regions of the image. *Front. Psychol.* **8**, 1718 (2017).

3. P. Neri, Object segmentation controls image reconstruction from natural scenes. *PLoS Biol.* **15**, e1002611 (2017).
4. B. A. Olshausen, D. J. Field, How close are we to understanding v1? *Neural Comput.* **17**, 1665–1699 (2005).
5. T. Sawada, A. A. Petrov, The divisive normalization model of V1 neurons: A comprehensive comparison of physiological data and model predictions. *J. Neurophysiol.* **118**, 3051–3091 (2017).

6. S. A. Cadena *et al.*, Deep convolutional models improve predictions of macaque V1 responses to natural images. *PLoS Comput. Biol.* **15**, e1006897 (2019).
7. W. Li, V. Piech, C. D. Gilbert, Contour saliency in primary visual cortex. *Neuron* **50**, 951–962 (2006).
8. C. D. Gilbert, W. Li, Top-down influences on visual processing. *Nat. Rev. Neurosci.* **14**, 350–363 (2013).
9. V. A. Lamme, The neurophysiology of figure-ground segregation in primary visual cortex. *J. Neurosci.* **15**, 1605–1615 (1995).
10. P. R. Roelfsema, F. P. de Lange, Early visual cortex as a multiscale cognitive blackboard. *Annu. Rev. Vis. Sci.* **2**, 131–151 (2016).
11. L. Kirchberger, S. Mukherjee, M. W. Self, P. R. Roelfsema, Contextual drive of neuronal responses in mouse V1 in the absence of feedforward input. *Sci. Adv.* **9**, eadd2498 (2023).
12. J. Poort, M. W. Self, B. van Vugt, H. Malkki, P. R. Roelfsema, Texture segregation causes early figure enhancement and later ground suppression in areas V1 and V4 of visual cortex. *Cereb. Cortex* **26**, 3964–3976 (2016).
13. J. Poort *et al.*, The role of attention in figure-ground segregation in areas V1 and V4 of the visual cortex. *Neuron* **75**, 143–156 (2012).
14. M. W. Self *et al.*, The segmentation of proto-objects in the monkey primary visual cortex. *Curr. Biol.* **29**, 1019–1029.e4 (2019).
15. P. Papale *et al.*, The representation of occluded image regions in area V1 of monkeys and humans. *Curr. Biol.* **33**, 3865–3871.e3 (2023).
16. M. Chen *et al.*, Incremental integration of global contours through interplay between visual cortical areas. *Neuron* **82**, 682–694 (2014).
17. P. Papale *et al.*, Foreground-background segmentation revealed during natural image viewing. *eNeuro* **5**, ENEURO.0075-18.2018 (2018).
18. W. Zuiderbaan, B. M. Harvey, S. O. Dumoulin, Image identification from brain activity using the population receptive field model. *PLoS One* **12**, e0183295 (2017).
19. P. Arbelaz, M. Maire, C. Fowlkes, J. Malik, Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**, 898–916 (2011).
20. V. A. Lamme, K. Zipser, H. Spekreijse, Figure-ground activity in primary visual cortex is suppressed by anesthesia. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 3263–3268 (1998).
21. S. O. Dumoulin, B. A. Wandell, Population receptive field estimates in human visual cortex. *Neuroimage* **39**, 647–660 (2008).
22. A. Kohn, R. Coen-Cagli, Multi-electrode recordings of anesthetized macaque V1 responses to static natural images and gratings. *CRCNS.org* (2015). <https://doi.org/10.6080/KOSB43P8>.
23. R. Coen-Cagli, A. Kohn, O. Schwartz, Flexible gating of contextual influences in natural vision. *Nat. Neurosci.* **18**, 1648–1655 (2015).
24. P. R. Roelfsema, P. S. Khayat, H. Spekreijse, Subtask sequencing in the primary visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5467–5472 (2003).
25. P. R. Roelfsema, Cortical algorithms for perceptual grouping. *Annu. Rev. Neurosci.* **29**, 203–227 (2006).
26. V. A. Lamme, P. R. Roelfsema, The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* **23**, 571–579 (2000).
27. J. Portilla, E. P. Simoncelli, A parametric texture model based on joint statistics of complex wavelet coefficients. *Int. J. Comput. Vis.* **40**, 49–70 (2000).
28. E. Y. Walker *et al.*, Inception loops discover what excites neurons most using deep predictive models. *Nat. Neurosci.* **22**, 2060–2065 (2019). [10.1038/s41593-019-0517-x](https://doi.org/10.1038/s41593-019-0517-x).
29. Y. Zhang, T. S. Lee, M. Li, F. Liu, S. Tang, Convolutional neural network models of V1 responses to complex patterns. *J. Comput. Neurosci.* **46**, 33–54 (2019).
30. T. Marques, M. Schrimpf, J. J. Dicarlo, Multi-scale hierarchical neural network models that bridge from single neurons in the primate primary visual cortex to object recognition behavior. *bioRxiv* [Preprint] (2021). <https://doi.org/10.1101/2021.03.01.433495> (Accessed 10 October 2024).
31. A. J. Keller, M. M. Roth, M. Scanziani, Feedback generates a second receptive field in neurons of the visual cortex. *Nature* **582**, 545–549 (2020).
32. W. S. Geisler, Visual perception and the statistical properties of natural scenes. *Annu. Rev. Psychol.* **59**, 167–192 (2008).
33. W. S. Geisler, J. S. Perry, B. J. Super, D. P. Gallogly, Edge co-occurrence in natural images predicts contour grouping performance. *Vision Res.* **41**, 711–724 (2001).
34. K. Friston, A theory of cortical responses. *Proc. Trans. R. Soc. B Biol. Sci.* **360**, 815–836 (2005).
35. J. R. Williford, R. von der Heydt, Figure-ground organization in visual cortex for natural scenes. *eNeuro* **3**, ENEURO.0127-16.2016 (2016).
36. H. Zhou, H. S. Friedman, R. von der Heydt, Coding of border ownership in monkey visual cortex. *J. Neurosci.* **20**, 6594–6611 (2000).
37. J. K. Hesse, D. Y. Tsao, Consistency of border-ownership cells across artificial stimuli, natural stimuli, and stimuli with ambiguous contours. *J. Neurosci.* **36**, 11338–11349 (2016).
38. C. C. Fowlkes, D. R. Martin, J. Malik, Local figure-ground cues are valid for natural images. *J. Vis.* **7**, 2 (2007).
39. D. G. Pelli, The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spat. Vis.* **10**, 437–442 (1997).
40. P. J. Bex, W. Makous, Spatial frequency, phase, and the contrast of natural images. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **19**, 1096–1106 (2002).
41. P. Bashivan, K. Kar, J. J. DiCarlo, Neural population control via deep image synthesis. *Science* **364**, eaav9436 (2019). [10.1126/science.aav9436](https://doi.org/10.1126/science.aav9436).
42. D. A. Klindt, A. S. Ecker, T. Euler, M. Bethge, "Neural system identification for large populations separating what and where" in *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)* (Curran Associates Inc., Red Hook, NY, 2017), pp. 3509–3519.
43. G. M. Boynton, J. B. Demb, G. H. Glover, D. J. Heeger, Neuronal basis of contrast discrimination. *Vision Res.* **39**, 257–269 (1999).
44. M. W. Self, R. N. Kooijmans, H. Super, V. A. Lamme, P. R. Roelfsema, Different glutamate receptors convey feedforward and recurrent processing in macaque V1. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 11031–11036 (2012).
45. H. Super, P. R. Roelfsema, Chronic multiunit recordings in behaving animals: Advantages and limitations. *Prog. Brain Res.* **147**, 263–282 (2005).
46. M. R. Cohen, J. H. Maunsell, Attention improves performance primarily by reducing interneuronal correlations. *Nat. Neurosci.* **12**, 1594–1600 (2009).
47. C. Palmer, S. Y. Cheng, E. Seidemann, Linking neuronal and behavioral performance in a reaction-time visual detection task. *J. Neurosci.* **27**, 8122–8137 (2007).
48. E. M. Trautmann *et al.*, Accurate estimation of neural population dynamics without spike sorting. *Neuron* **103**, 292–308.e4 (2019).
49. B. C. Motter, Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *J. Neurophysiol.* **70**, 909–919 (1993).
50. D. H. Brainard, The psychophysics toolbox. *Spat. Vis.* **10**, 433–436 (1997).
51. O. Nestares, D. J. Heeger, Robust multiresolution alignment of MRI brain volumes. *Magn. Reson. Med.* **43**, 705–715 (2000).
52. S. M. Smith *et al.*, Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* **23**, S208–S219 (2004).
53. P. A. Yushkevich *et al.*, User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* **31**, 1116–1128 (2006).
54. B. A. Wandell, S. Chial, B. T. Backus, Visualization and measurement of the cortical surface. *J. Cogn. Neurosci.* **12**, 739–752 (2000).
55. B. M. Harvey, S. O. Dumoulin, The relationship between cortical magnification factor and population receptive field size in human visual cortex: Constancies in cortical architecture. *J. Neurosci.* **31**, 13604–13612 (2011).
56. B. A. Wandell, S. Chial, B. T. Backus, Visualization and measurement of the cortical surface. *J. Cogn. Neurosci.* **12**, 739–752 (2000).
57. B. A. Wandell, S. O. Dumoulin, A. A. Brewer, Visual field maps in human cortex. *Neuron* **56**, 366–383 (2007).
58. E. A. DeYoe *et al.*, Mapping striate and extrastriate visual areas in human cerebral cortex. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 2382–2386 (1996).
59. S. A. Engel, G. H. Glover, B. A. Wandell, Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb. Cortex* **7**, 181–192 (1997).
60. M. I. Sereno *et al.*, Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* **268**, 889–893 (1995).
61. K. J. Friston *et al.*, Analysis of fMRI time-series revisited. *Neuroimage* **2**, 45–53 (1995).
62. K. J. Friston *et al.*, Event-related fMRI: Characterizing differential responses. *Neuroimage* **7**, 30–40 (1998).
63. W. Zuiderbaan, B. M. Harvey, S. O. Dumoulin, Modeling center-surround configurations in population receptive fields using fMRI. *J. Vis.* **12**, 10 (2012).
64. S. O. Dumoulin, T. Knapen, How visual cortical organization is altered by ophthalmologic and neurologic disorders. *Annu. Rev. Vis. Sci.* **4**, 357–379 (2018).
65. J. Freeman, E. P. Simoncelli, Metamers of the ventral stream. *Nat. Neurosci.* **14**, 1195–1201 (2011).
66. P. Papale *et al.*, V1 multi-unit activity in response to object and non-object borders. Open Science Foundation. <https://osf.io/qpc2d/>. Deposited 10 October 2024.