The influence of objecthood on the representation of natural images in the visual cortex

Paolo Papale[1,2,‡], Wietske Zuiderbaan[3,‡], Rob. R.M. Teeuwen[1], Amparo Gilhuis[1], Matthew W. Self[1,¶],

Pieter R. Roelfsema[1,4,5,¶,*] and Serge O. Dumoulin[1,3,6,7,¶]

1. Department of Vision & Cognition, Netherlands Institute for Neuroscience (KNAW), 1105 BA Amsterdam, Netherlands

2. Momilab Research Unit, IMT School for Advanced Studies Lucca, 55100 Lucca, Italy

3. Spinoza Centre for Neuroimaging, 1105 BK Amsterdam, Netherlands

4. Department of Integrative Neurophysiology, VU University, 1081 HV Amsterdam, Netherlands

5. Department of Psychiatry, Academic Medical Centre, Postbus 22660, 1100 DD Amsterdam, Netherlands.

6. Department of Experimental and Applied Psychology, VU University Amsterdam, Amsterdam 1181 BT, Netherlands

7. Department of Experimental Psychology, Helmholtz Institute, Utrecht University, 3584 CS Utrecht, Netherlands

‡ shared first author contribution

¶ shared senior author contribution

* correspondence should be addressed to: *p.roelfsema@nin.knaw.nl*

1

Abstract

Neurons in early visual cortex are not only sensitive to the image elements in their receptive field but also to the context determining whether the elements are part of an object or background. We here assessed the effect of objecthood in natural images on neuronal activity in early visual cortex, with fMRI in humans and electrophysiology in monkeys. We report that boundaries and interiors of objects elicit more activity than the background. Boundary effects occur remarkably early, implying that visual cortical neurons are tuned to features characterizing object boundaries in natural images. When a new image is presented the influence of the object interiors on neuronal activity occurs during a late phase of neuronal response and earlier when eye movements shift the image representation, implying that object representations are remapped across eye-movements. Our results reveal how object perception shapes the representation of natural images in early visual cortex.

Introduction

The visual scenes that we perceive are filled with objects. We readily identify the extent of the objects and their boundaries, a perceptual organization process that is important for our understanding of an image's meaning. Accordingly, the judgments of people who are asked to mark regions occupied by objects and their boundaries are highly consistent[1]. Object and boundary perception even influence low-level vision, because image elements at object boundaries are better perceived than image elements at less relevant image locations[2,3]. Furthermore, image elements of objects have a higher perceived contrast than those that are part of the background[4]. Despite these influences on low-level visual perception, it is not yet well understood how objecthood influences neuronal representations in early visual cortex[5].

2

45 Classical descriptions of the activity of neurons at the early levels of the visual system focus on the

46 features that drive neurons, such as the contrast and orientation in a neuron's receptive field (RF).

47 In addition, there are also non-classical, contextual influences on neuronal activity, which originate

48 from outside the neurons' RFs and play a role in the grouping of features into objects. Here we focus

49 on two such effects: boundary modulation (BoM) related to the detection of object boundaries, and

50 object-background modulation (OBM) related to the grouping of object features into objects and

51 their segregation from the background.

52 Neurons in the primary visual cortex (V1) and area V4 increase their firing rate when their RF is

53 centered on an elongated contour that extends well beyond their RF (Fig. 1a)[6,7]. Elongated contours

54 are relevant for perceptual organization because they usually signal the borders of objects in natural

55 scenes, whereas shorter contours are more likely to be part of the background[8,9]. BoM is the extra

56 activity elicited by contours that demarcate object boundaries (Fig. 1a,d). Similarly, V1 and V4

57 neurons exhibit stronger responses when their RF falls on the interior of a perceptual object than

58 when it falls on the background (Fig. 1b)[10–12]. This OBM occurs for all image regions that are part of

59 an object, suggesting that the response enhancement could cause the binding of the distributed

60 representation of features in early visual cortex into coherent perceptual objects[13]. This view is

61 supported by the finding that objects relevant for behavior elicit stronger OBM than objects that

62 are not, implying a relation between OBM and object-based attention that co-selects all features of

63 a relevant object[11]. BoM and OBM are thought to reflect the recurrent interactions within and across

64 visual areas[14] that determine the spread of enhanced neuronal activity and thereby the perception

65 of spatially extended objects in a scene[13].

66 So far, BoM and OBM have only been measured with artificial stimuli, such as textures and displays

67 with many line elements (Fig. 1a-c). Establishing the relevance of these signals for natural vision is

3

68   challenging, yet important, because neuronal response properties that do not play a role in the

69   perception of natural stimuli are likely to be of limited relevance[15]. A recent study explored

70   contextual signals in V1 elicited by more complex shapes, such as the texture-defined 'U' of Figure

71   1c[16], but researchers have, to our knowledge, not yet examined BoM and OBM with natural visual

72   stimuli. If BoM occurs in natural images, we predict that the more salient object boundaries elicit

73   stronger neuronal activity than image elements of the background. Similarly, if OBM occurs for

74   natural images, extra activity should be elicited by object interiors compared to the background.

75   To investigate the influence of objecthood on neuronal activity in early visual cortex, we used the

76   Berkeley Segmentation data set (BSD), a library of natural images in which human observers marked

77   object boundaries[1]. We used functional MRI to examine neuronal responses across many regions of

78   visual cortex in humans and we also recorded multi-unit activity in V1 and V4 of monkeys to gain

79   insight into the temporal profile of spiking activity. We report that objecthood influences neuronal

80   activity. Object boundaries increased the early neuronal responses and object interiors enhanced

81   activity during a later phase of the response. When subjects made eye movements across the

82   images, these contextual effects carried over from one fixation to the next, implying that objects

83   are remapped across eye movements in early visual cortex[17,18].

84
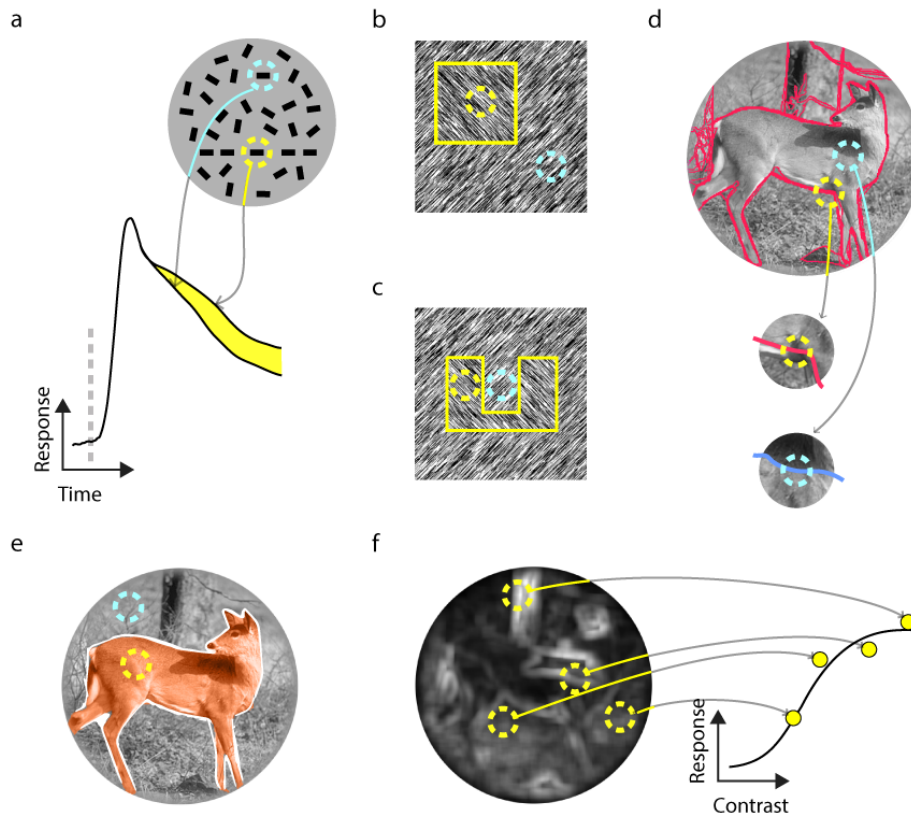
Figure 1. The influence of object perception on neuronal responses in early visual cortex.
a, The response of V1 and V4 neurons is enhanced when their RF falls on an elongated contour that extends well beyond their RF[6]. **b,c,** V1 and V4 neurons exhibit stronger responses when their RF falls on the interior of a perceptual object (square with different orientation in b and 'u' shape in c) than when it falls on the background[16]. **d,e**, We ask whether differences between object borders (yellow in d) and other image regions (cyan in d) and between the interiors of objects (yellow in e) and the background (cyan in e) in natural images influence the response of visual cortical neurons. **f**, We compared the response amplitudes evoked by image elements of objects and the background, taking the local contrast in the (p)RF into account.

Results

Objecthood modulates responses in human early visual cortex

In the first experiment, we used ultra-high field fMRI at 7 Tesla in four human participants to investigate OBM and BoM within natural images (Fig. 2). Our analysis separated BoM (Fig. 1d) and OBM (Fig. 1e) from the influence of the contrast of image elements by evaluating the influence of image properties in the population receptive field (pRF) on the neuronal responses at each cortical

5

102 location. We chose 45 images from the BSD (Fig. S1) for which the object boundaries had been

103 annotated by an independent group of human observers[1].

104 To separate the influence of image contrast from object perception we computed the contrast

105 response functions (CRFs; Fig. 1f) for six regions of interest: V1, V2, V3, human V4 (hV4), the lateral

106 occipital visual field maps 1/2 (LO-1/2) and V3-a/b. First, in a separate experiment, we measured

107 the population receptive field for each cortical location (pRF; Fig. S2) [19]. Second, to compute the CRF,

108 we estimated the response amplitude as function the root mean square (RMS) contrast in each

109 pRF[20] (10 contrast bins, Fig. 1f).

110 Next, we computed the CRF separately when the pRF fell on an object border versus a non-border

111 image region, and when it fell an object region versus on the background for every image (Fig.

112 2b)[1,21]. We defined object borders in the BSD images as those that were frequently marked by the

113 observers as object boundaries, and contrasted them to non-border image regions that were not

114 marked (Fig S3). These non-border regions could be part of the object interior (as the example in

115 Fig. 1d and Fig. S3) or background (see Methods).

116 The computation of the CRF allowed us to separate BoM and OBM from the influence of contrast.

117 Cortical responses elicited by object borders were significantly higher than those elicited by non-

118 border image regions in areas V1, V2 and V3 (Fig. 2c; all ps < 0.001, bootstrap test, see Methods),

119 but not in areas V3ab, hV4 and LO-1 and 2 (Fig. S2). Thus, we observed significant BoM in V1-V3.

120 Object boundaries of a particular contrast elicit a larger response, on average, than image regions

121 with the same contrast that do not coincide with object boundaries.

122 To examine the influence of OBM, we compared CRFs of cortical locations with pRFs on object versus

123 background regions (Fig S3). The response amplitude when a pRF was centered on an object was

6

124 significantly stronger than when it was centered on the background in V1 and V2 (Fig. 2d; all ps <

125 0.001, bootstrap test) but not in V3, hV4, LO-1/2 and V3A/b. We observed the same pattern of

126 results also at the level of individual participants, for both BoM and OBM (Figure S4).

127 We conclude that object borders elicit larger response amplitudes in early visual cortex than non-

128 border image regions, and that object regions elicit more activity than background regions, even if

129 image contrast is the same.

130
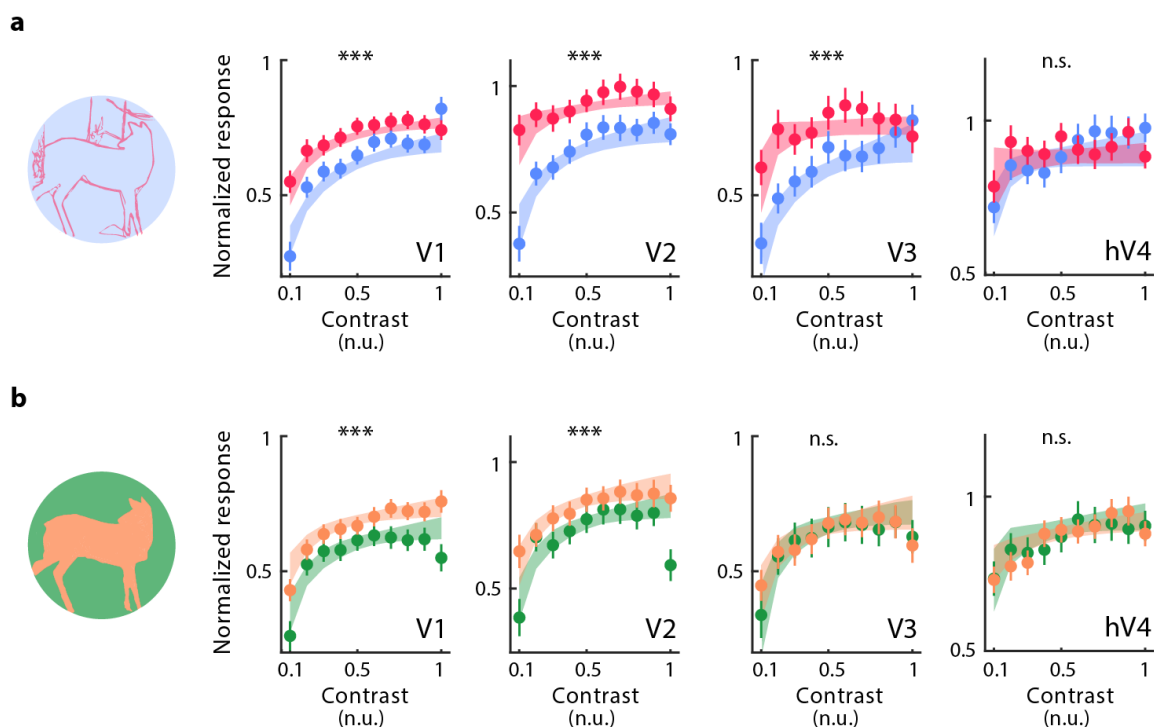
131

132 **Figure 2. Objecthood influences responses in human early visual cortex.**
133 Four participants viewed 45 natural images while their brain responses were recorded with fMRI. **a**,
134 Response amplitudes elicited by object boundaries (red) and background contours (blue) as a
135 function of contrast in the pRF (x axis). **b**, Response amplitudes elicited by object interiors (orange)
136 and the background (green). fMRI responses were normalized to the response to a full-field, 100%
137 contrast stimulus. Shaded regions denote 95% confidence intervals determined by bootstrapping.
138 Bars represent SEM across images (*** indicates p < 0.001, bootstrap test; n.s. non-significant).
139

140   *Objects and their boundaries enhance the spiking activity of V1 and V4 neurons but at different*

141   *latencies*

142   A limitation of fMRI is its poor temporal resolution and its indirect relation to spiking activity[22].

143   Therefore, we recorded spiking activity with chronically implanted electrode arrays elicited by BSD

144   stimuli in two macaque monkeys. We placed the arrays in areas V1 and V4 and recorded multi-unit

145   spiking activity (MUA). Whereas the pRFs in the MRI experiment covered the entire images, the RFs

146   of the V1 and V4 neurons in the electrophysiology experiment were confined to a limited region of

147   the visual field. To increase the sample of image patches falling in the RFs, we trained the monkeys

148   to fixate at multiple locations on a total of four images (Fig. 3a).

149   At the start of the trial, the monkey directed gaze to a fixation point on a gray background. We

150   presented the image once the monkey had maintained fixation for 300ms. After a delay of 400ms,

151   we presented a new fixation point and the monkey made a saccade to it and maintained fixation for

152   a further 400ms (Fig. 3a). The repositioning of gaze was repeated for a total of 6 positions (sampled

153   from a uniformly spaced grid with ~500 points) per trial. The neuronal response elicited by the image

154   appearing at the first gaze position differs from that for later fixations because the image suddenly

155   appears in the RF. Later fixations are preceded by saccades causing a rapid movement of part of the

156   image through the RF. Furthermore, the image is now familiar and the monkeys may have

157   recognized and segmented the objects during previous fixations. We therefore separately analyzed

158   the response elicited by stimulus onset (Fixation 1) and later fixations (Fixations 2-6). The results for

159   fixations 2-6 were comparable and we therefore pooled the data across these fixations (Fig. S5).

160

8
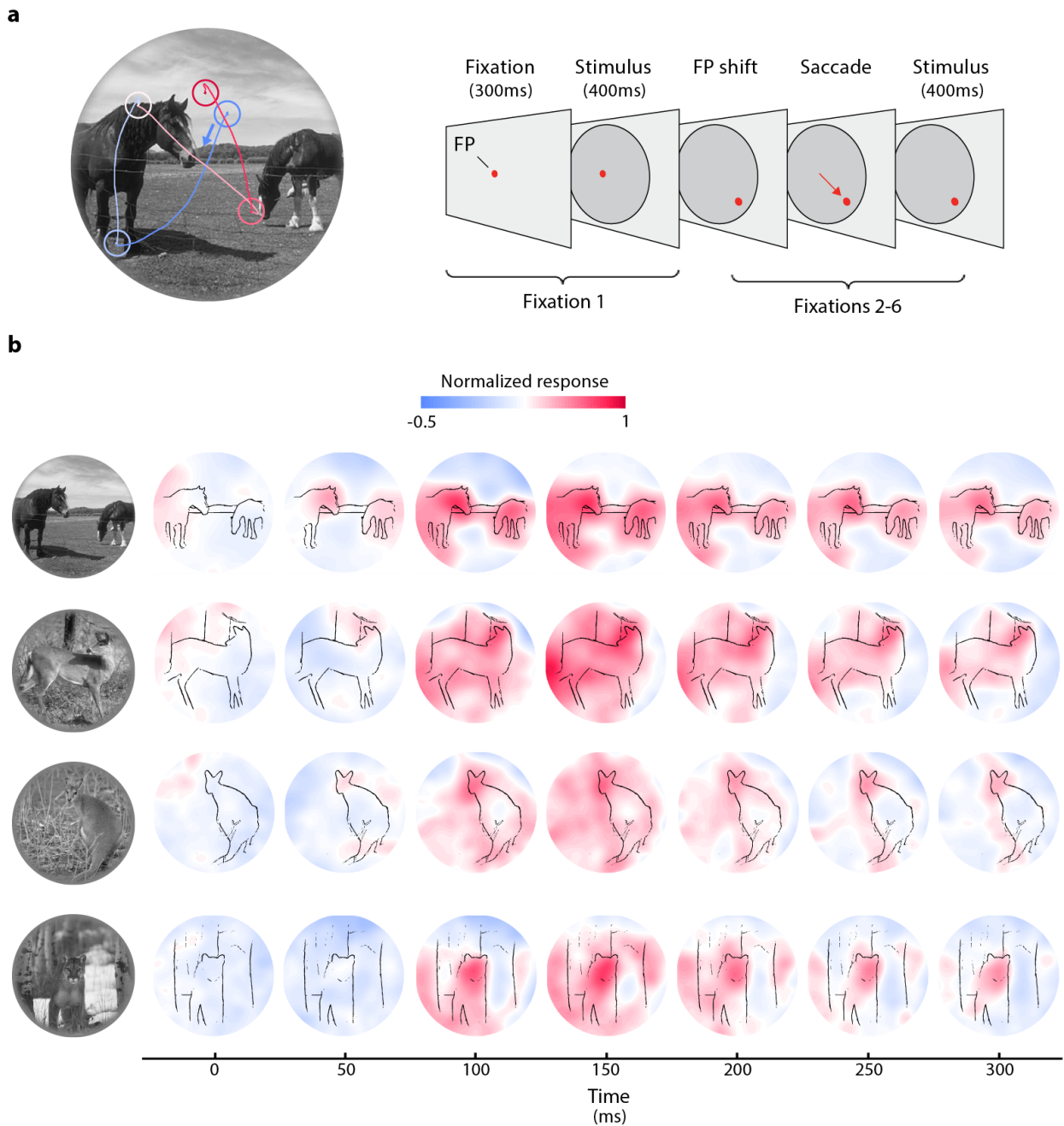
Figure 3. Influence of objects on spiking activity in areas V1 and V4 evoked by natural scenes.
a, The monkey performed a sequence of eye-movements across the natural images. We presented a natural image once the monkey had maintained gaze on a red fixation point for 300ms. After a delay of 400ms, a new fixation point appeared and the monkey made a saccade to it and maintained fixation for a further 400ms. Per trial, the monkey made a total of 5 eye movements to fixation points sampled from a uniformly spaced grid (~500 points). **b**, Overlay of V4 spiking activity over the natural images at different time points (average of fixations 2-6). V4 response is determined by contrast, local and global image structure and these factors were disentangled in subsequent analyses. Fig. S3b shows the same analysis for fixation 1 in V4 and fixation 1 and 2-6 in V1.

9

172   We first examined the overall level of activity in V1 and V4 elicited by the four natural images (Fig.

173   3b and S3). The response profiles suggest that extra activity is focused on the objects, but the

174   influences of objecthood and contrast were not yet separated in this analysis. To disentangle the

175   influence of BoM and OBM from that of contrast, we determined CRFs by binning the contrast in

176   the RF of neurons in area V1 (77 recording sites, 44 in monkey B and 33 in monkey M) and V4 (22

177   sites in monkey B), separately for contours that demarcate object boundaries and those that do not.

178   Object borders elicited stronger spiking activity (time-window 0-300ms) than non-object image

179   regions with the same contrast in V1 and V4 (Fig. 4a). BoM occurred during the first fixation as well

180   as during later fixations (all ps < 0.001, bootstrap test) and was present at many V1 recording sites

181   in monkey B (fixation 1, 66% of the sites; fixation 2-6, 77%; Fig. S6) and monkey M (fixation 1, 45%;

182   fixation 2-6, 51%) and at V4 recording sites as well (monkey B, fixation 1, 45%; fixation 2-6, 45%).

183   We determined BoM latency by fitting a curve to the difference in activity elicited by the object

184   borders and non-border image regions, averaged across contrast bins. We estimated latency as the

185   time-point at which the fitted function reached the 33% of its maximum (see Methods)[11,12,23]. In V1,

186   the BoM latency was 50ms in both fixation 1 and in later fixations. BoM latency was not significantly

187   different from the latency of the visually driven response (49ms for fixation 1 and 29ms for later

188   fixations), neither for fixation 1 nor for the later fixations (both ps > 0.05, bootstrap test). The same

189   was true for V4 in both conditions (BoM: 59ms for fixation 1; 49ms for fixations 2-6; onset of

190   response: 61ms for fixation 1; 54ms for fixations 2-6; both ps > 0.05, bootstrap test). We cannot

191   directly compare latencies between fixation 1 and later fixations, because in fixation 1 the image

192   replaced a grey background whereas the image moved through the receptive fields preceding the

193   later fixations. We corrected for this difference by computing $Lat_{BoM\text{-}Vis}$, the difference between the

194   BoM latency and the visual latency and compared it between fixation 1 and later fixations across

195 recording sites (Fig. S7). $Lat_{BoM-Vis}$ did not differ between fixation 1 and fixations 2-6 (p > 0.05,

196 Wilcoxon signed-rank test; Fig. S7).

197 Next, we compared the response elicited by object interiors to that elicited by background regions

198 (Fig. 4b). Regions that were part of objects elicited more activity in V1 and V4 than background

199 regions, both during the first fixation and later fixations (all ps < 0.001, bootstrap test). Many V1

200 recording sites exhibited OBM (monkey B, fixation 1, 41%, fixations 2-6, 52%; monkey M, fixation 1,

201 67%, fixations 2-6, 36% of sites with p<0.05, bootstrap test) and OBM was also present in V4

202 (monkey B, fixation 1, 41% of sites in V4, fixations 2-6, 45%). Hence, OBM also occurs for natural

203 images: image elements of objects elicit a stronger activity than those that are part of the
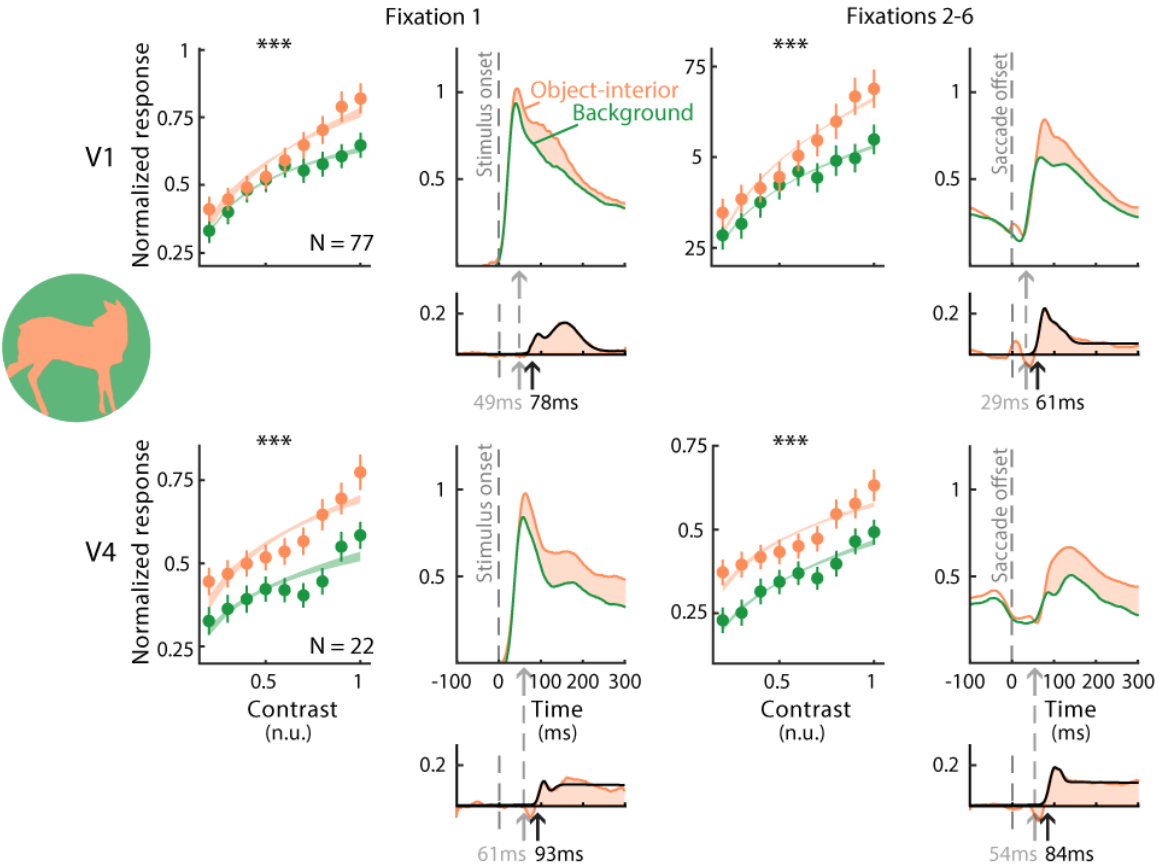
204 background.

205 In V1, the latency of OBM during the first fixation was 78ms, which was later than the onset of the

206 visually driven response (p < 0.05, bootstrap test). We next examined whether the OBM latency was

207 shorter for later fixations, because the monkeys may have segmented the image in figure and

208 background during the previous fixations. Interestingly, the median OBM latency across sites for

209 fixations 2-6 was only 61ms, and not significantly different from the visually driven response and

210 BoM (p > 0.05, bootstrap test). To correct for the difference in visual stimulation we computed

211 $Lat_{OBM-Vis}$, the difference between the latency of OBM and the visual response across recording sites.

212 In V1, $Lat_{OBM-Vis}$ was 13ms shorter during fixations 2-6 than during fixation 1 (p < 0.05, Wilcoxon

213 signed-rank test; Fig. S7). In V4, the latency of OBM was later than the onset of visually driven

214 response both for fixation 1 and fixations 2-6 (both ps < 0.05, bootstrap test). The difference in

215 $Lat_{OBM-Vis}$ between the fixation periods was not significant, but we cannot exclude the possibility that

216 this was caused by the smaller number of V4 recording sites.

11

217   The earlier OBM in V1 during fixations 2-6 suggests that it may have carried over from earlier

218   fixations during which the monkeys had already recognized and segmented the objects, in

219   accordance with previous studies demonstrating that image segmentation results can be remapped

220   across saccades[17,18].
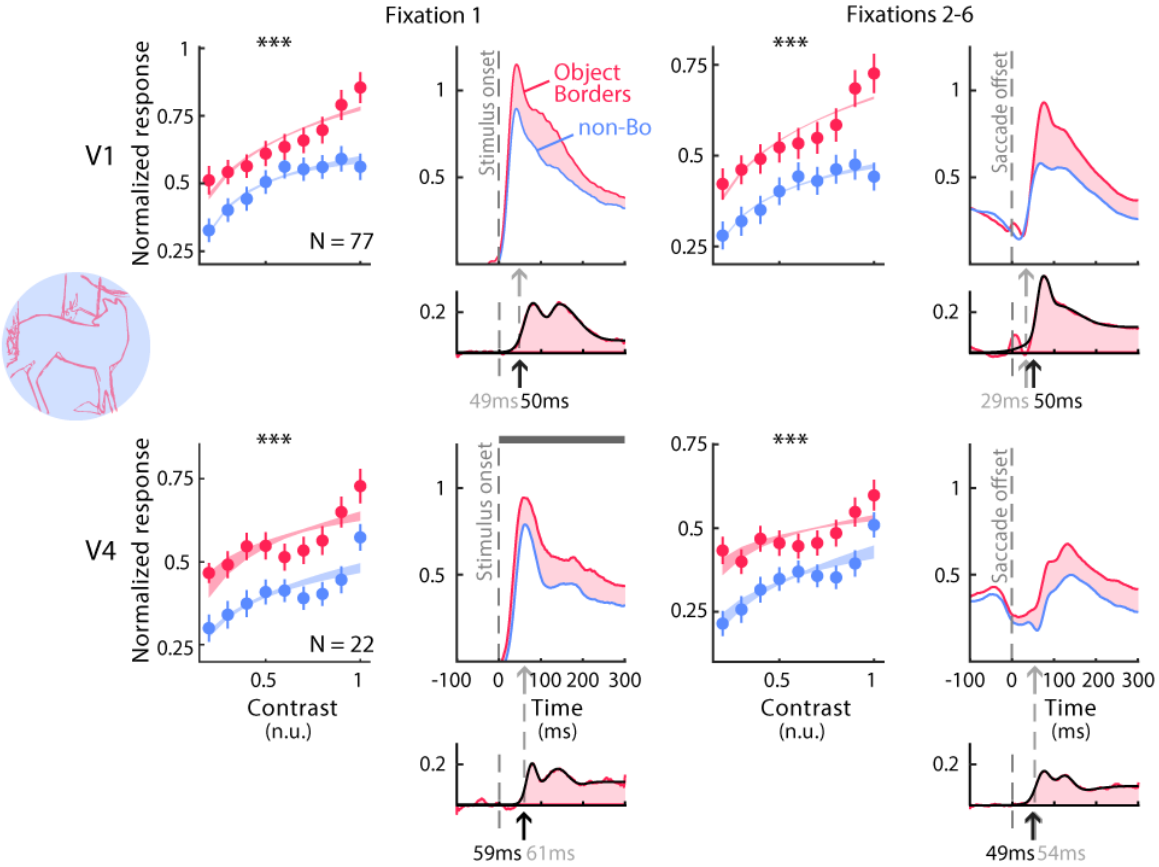
222 **Figure 4. Object borders and their interiors enhance the spiking activity of V1 and V4 neurons**
223 **relative to background regions.**
224 **a**, Average CRFs (left) and MUA time-courses (averaged across contrast bins; right) in V1 (top row)
225 and V4 (bottom row) for object-borders (red) and non-border image regions (blue). BoM is
226 significant in both areas. Shaded regions around the CRFs denote 95% confidence intervals
227 (determined by bootstrapping). Error bars indicate SEM across recording sites (***, p < 0.001,
228 bootstrap test). **b**, CRFs and MUA time-course elicited by the object-interior (orange) and
229 background (green). Black arrows indicate the latency of either BoM/OBM and gray arrows the
230 latency of the visually driven response.

231

232 BoM entails a comparison of the response elicited by object borders and other image elements,

233 which were inside the objects or in the background, whereas OBM entails a comparison of image

234 elements inside objects to background elements (Fig S3). BoM and OBM are not independent

235 because both measures compare neuronal activity evoked by some of the object elements and

236 background elements. We therefore also investigated the amount of unique variance in the activity

237 of V1 and V4 neurons (time window 0-300ms) explained by BoM, OBM and contrast (Fig. 5a; see

238 Methods). Each predictor explained a significant amount of unique variance in both areas and both

239 fixation conditions (all ps < 0.001, t-test). In V1, contrast explained 41.0% of the variance, BoM 8.3%

240 and OBM 5.5% during the first fixation and the values increased slightly to 44.8%, 10.1% and 6.4%

241 for fixations 2-6, respectively. In V4, contrast explained 19.0%, BoM 21.4% and OBM 16.0% of the

242 variance during the first fixation and these values were 20.8%, 21.1% and 35.0% for the later

243 fixations. Hence, BoM and OBM accounted for a significant fraction of the explainable variance.

244 Contrast explained less variance in V4 than in V1 whereas the contributions of BoM and OBM were

245 larger in V4. It is of interest that the fraction of variance explained by OBM in V4 increased from

246 16% for the first fixation to 35% for the later fixations. This result suggests that the extra activity

247 elicited by the interior of objects builds up across between the first presentation of the image and

248 subsequent eye movements, possibly because the scene is already known.

249

250   *The tuning of the early V1 response is selective for object borders*

251   We were surprised to find BoM in V1 at a latency of 50ms (Fig. 4a), because it is much earlier than

252   the latency of ~95ms typically observed for elongated contours in synthetic images[7]. The longer

253   BoM latency of previous studies is compatible with an effect of feedback from higher cortical areas

254   to V1, but a latency of 50ms might be too short for such a feedback loop. An important difference

255   between the present approach and previous studies with synthetic images is that we did not equate

256   the features of contours that form the boundaries of objects and those that were in the background,

257   even though contrasts were matched. We therefore hypothesized that object contours in natural

258   images have other features, on average, than background contours[9,24], which could explain the early

259   BoM. In other words, some V1 neurons might be tuned to the features of object contours and

260   extract them in their feedforward response, driven from within the RF.

261   We exploited recent advances in artificial neural networks (ANNs) to study the tuning of V1

262   recording sites and to examine if it can account for the extra activity elicited by object boundaries[25–

263   27]. As a model for V1 tuning we chose layer conv3_1 of VGG-19 (and several other models, Fig. S8),

264   which is the state of the art in predicting V1 responses to natural images[28,29], and used a two-stage

265   convolutional mapping to take both the spatial and feature selectivity of neurons at individual

266   recording sites into account (see Methods)[26]. We confirmed previous studies[28,29] demonstrating

267   that the ANN approach for the modeling of V1 tuning outperforms previous models (Fig. S8). To gain

268   insight into the tuning of the V1 neurons and visualize their preferred features, according to the

269   ANN model, we examined the synthetic images that maximized the model response. Figure 5c

270   illustrates a few of these synthetic images (for illustration purposes; we did not present these

271   images as stimuli). We then applied these RF models to an independent set of 100 natural images

272   that had been annotated by human observers to examine if they indeed predicted extra activity for

15

273   object boundaries (Fig. 5b,c and S9). Specifically, we filtered the unseen images with the RF models

274   (step 1 in Fig. 5b) and compared the filtered images to the human annotations (step 2 in Fig. 5b).

275   For every recording site we determined their border detection performance (BoP), a measure that

276   quantifies how well the V1 RF models predict the human-annotated borders (step 3 in Fig. 5b and

277   Y-axis in Fig. 5c). BoP is an accuracy score that takes the uneven class distribution of salient and non-

278   salient borders into account (F-measure, see Methods). Interestingly, 93% of the VGG-19 models of

279   V1 tuning detected objected contours above chance level (ps < 0.05, permutation test), which

280   indicates that V1 tuning is indeed useful for boundary detection.

281   What is the relation between BoM elicited in V1 by the four pictures of our electrophysiological

282   experiments and the BoP of the same recording sites for a different set of images? We computed

283   the correlation between BoM in the early time-window (25-75ms: x-axis in Fig. 5c) and BoP (y-axis

284   in Fig. 5c), across recording sites. The correlation coefficient was 0.25 (p = 0.037, t-test), indicating

285   that V1 neurons that express BoM at an early latency are tuned to low-level feature differences that

286   discriminate between object and non-object contours (Fig. 5c).

287   We next examined how much information about object contours is present across the recorded

288   population of V1 neurons. We built a binary classifier based on the early (25-75ms) V1 activity in

289   response to trials with object-contours or elements with the same contrast in the neurons' RF. To

290   ensure that complex patterns signaling the presence of an object (e.g., the entire head of an animal)

291   could not be detected by the RFs, we only included the 19 recording sites with smallest RFs (<1.5°)

292   in this analysis. Classification of object boundaries during single fixations had an accuracy of 73.5%,

293   which is well above the chance level of 50% (Fig. 5d, top, red bar; p<0.001, bootstrap test).

294   Interestingly, when we used the activity of the entire conv3_1 layer of the VVG-19 ANN to detect

295   object-borders the accuracy was similar (Fig. 5d, bottom, red bar; 66.4%, p<0.001, bootstrap test).

16

296 Hence, object-contours can be detected with a reasonable accuracy based on local information in

297 individual RFs.

298 We carried out an extra experiment to confirm that the early BoM reflects V1 tuning rather than a

299 contextual influence. We removed the context by copying circular image patches from the BSD that

300 matched the V1 MUA RFs in size onto a grey background. We chose patches with a similar RMS

301 contrast that did or did not contain an object border and centered them on the RFs of neurons at

302 50 recording sites in monkey B (Fig. 5e). As predicted, patches with object borders elicited a slightly

303 stronger V1 response than patches without object borders with the same contrast ($p < 0.001$,

304 Wilcoxon signed rank test; Fig. 5f,g). Hence, the tuning of V1 neurons indeed explains a fraction of

305 the extra activity elicited by object boundaries.

306 Our finding that BoM is partially explained by V1 tuning begs the question of a possible contribution

307 of V1 tuning to OBM, i.e. the extra activity by the object interior. We therefore also examined low-

308 level differences between image elements of objects and backgrounds and built a binary (linear)

309 classifier to discriminate between object and background regions, based on the early (25-75ms)

310 response of the same 19 V1 recording sites as above, using trials with the same contrast. The

311 classification accuracy during single fixations was 69.9% (Fig. 5d, top, orange bar; $p<0.001$, bootstrap

312 test) and it was in the same range for the conv3_1 layer of VVG-19 (Fig. 5d, bottom, orange bar;

313 74.0%, $p<0.001$, bootstrap test). Thus, even though OBM emerges later than BoM, the activity of a

314 small number of V1 neurons is enough to differentiate between features that characterize the

315 interior of objects and the background.

316

17
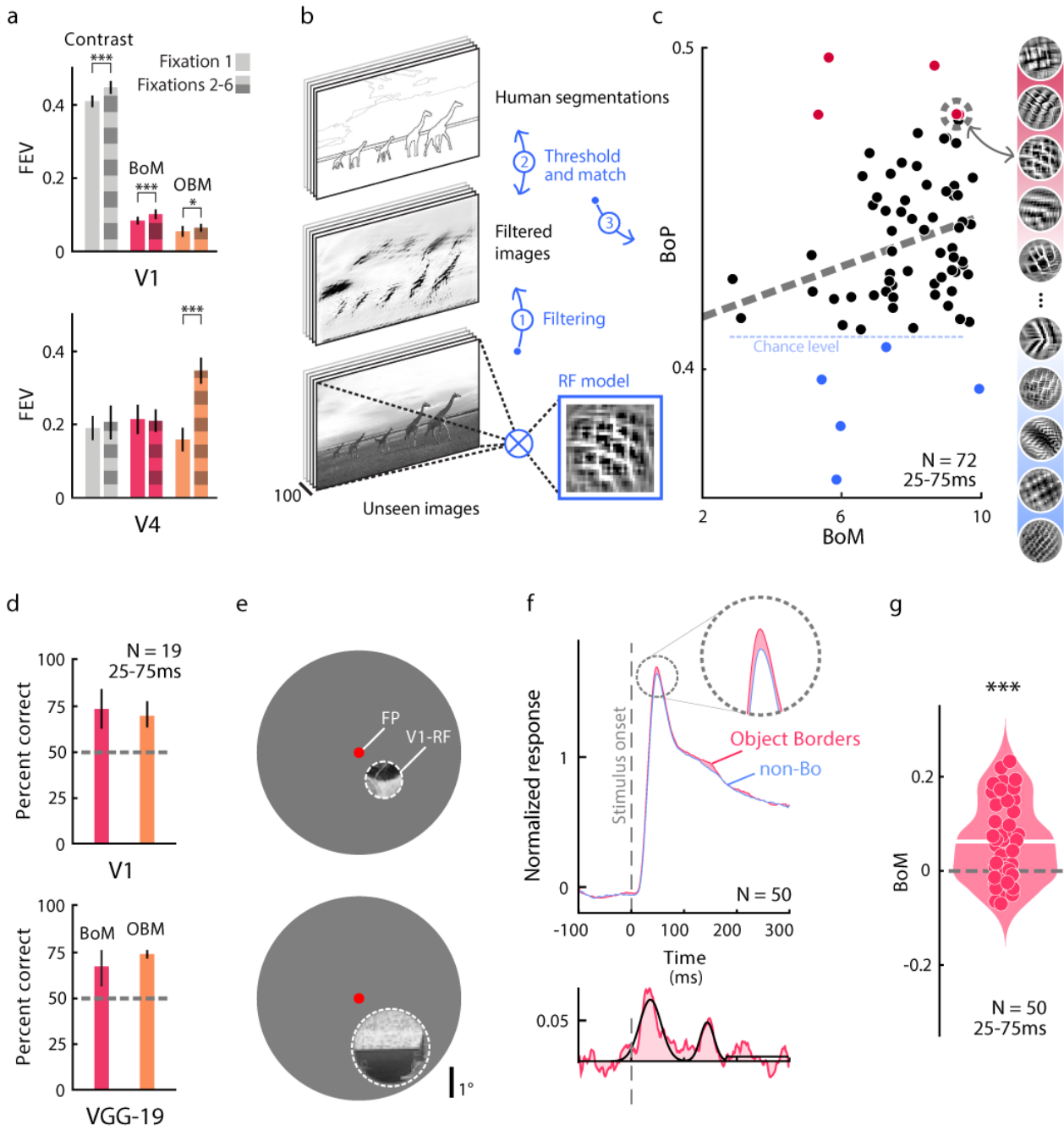
Figure 5. Explained variance in V1 and V4 and V1 tuning during the onset response.
a, Fraction of variance of V1 and V4 activity explained by contrast, BoM and OBM (0-300ms time window) for fixation 1 and fixations 2-6 (***: $p < 0.001$; *: $p < 0.05$, Wilcoxon signed rank). Error bars denote SEM. **b**, We derived RF models for each V1 recording site using ANNs, and applied them to a separate annotated set of natural images to examine how well they can detect object borders. We calculated a measure of border-detection performance (BoP) for every V1 recording site. Step 1, applying the V1 tuning to the image. Step 2, thresholding of activity and correlation with human judgements. Step 3, measurement of BoP of the V1 recording site. **c**, Correlation between BoM (x-axis) and BoP (y-axis) across V1 recording sites ($p < 0.05$, parametric test). Blue dashed line, significance threshold for border detection ($p < 0.05$, permutation test). **d**, Accuracy of binary

18

328 classifiers of object contours (red bar) and interiors (orange bar) based on the early response (25-
329 75 ms) of 19 V1 sites (upper panel) or the VGG-19 conv3_1 layer (lower panel). Classifiers detected
330 object contours and interiors above chance level. Error bars denote 95% confidence intervals
331 (determined by bootstrapping). **e**, Example isolated BSD image patches matching RFs of different
332 V1 recording sites. **f**, Time-course of the V1 responses. Object borders elicited stronger early activity
333 than non-border image patches. **g**, Distribution of early (25-75ms) BoM elicited by image patches
334 across recording sites (white bar indicate the median BoM; ***, p < 0.001, Wilcoxon signed-rank
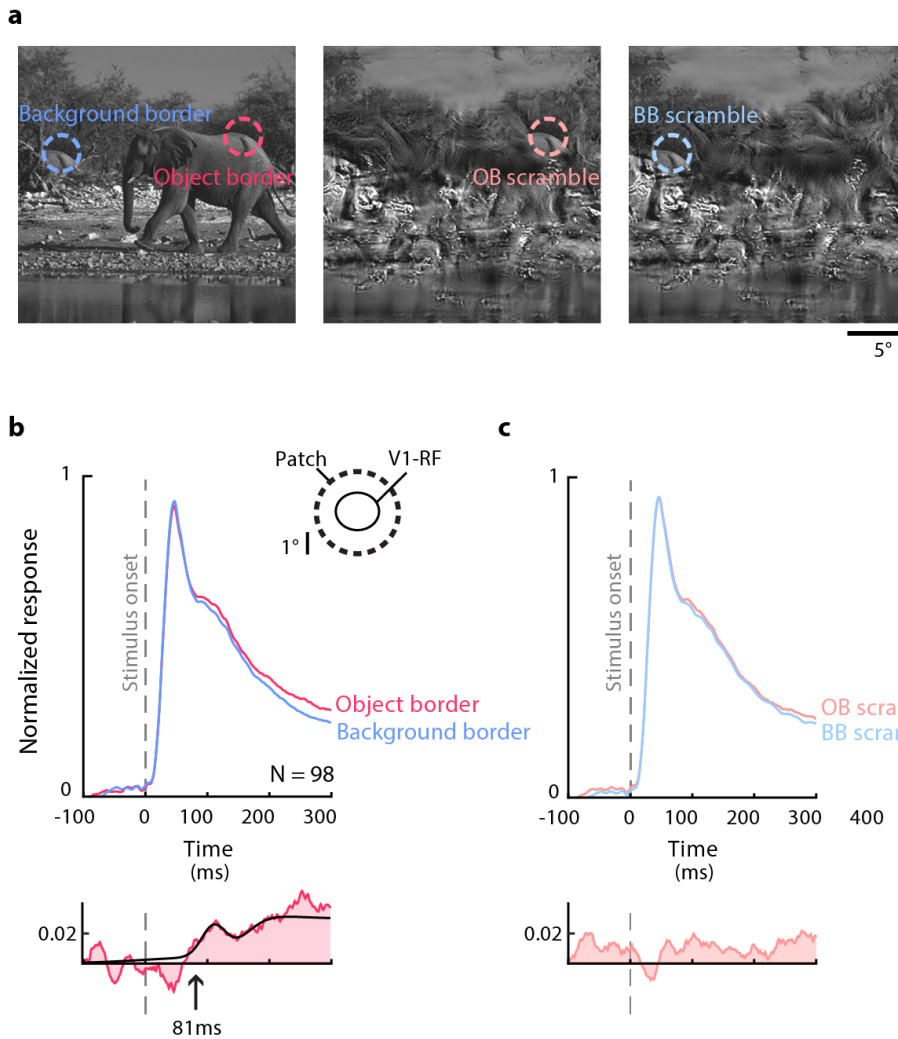335 test).

336

337 *Contextual BoM in natural images*

338 The early BoM in natural images is driven by the information in the RF. It differs from BoM in

339 previous studies[6,7], in which it was a contextual effect driven by information outside the neurons'

340 RF. Does BoM also occur for natural images if the image elements in the RFs are kept the same? In

341 a further experiment, we placed the RF of 98 V1 recording sites (68 in monkey B and 30 in monkey

342 M) on object borders and other locations in 12 natural images from the BSD, while keeping the

343 image patch in the RF constant (Fig. 6a). Specifically, we copied an image patch with an object border

344 and pasted it at a background location to create a condition in which the same image patch is not

345 perceived as object border. An example image is shown in Fig. 6a (left panel) where we copied a

346 part of the back of the elephant into the background. On average, the object contours elicited a

347 stronger V1 response than the same image patches presented at background locations (p < 0.001,

348 Wilcoxon signed-rank test across recording sites; Fig. 6b). The latency of BoM in this experiment was

349 81ms, i.e. it now occurred during the delayed phase of the V1 response.

350 As a control, we placed the image patch at identical locations of synthetic metamers of these

351 images. The metamers had the same orientations, phases, spatial frequencies, auto- and cross-

352 correlations and marginal statistics, but the layout of objects was scrambled[30]. In the example

353 metamer of Fig. 6a (middle and right panels), the transitions between water, trees and air were at

354 the same locations but the elephant was removed (other example metamers are shown in Fig. S10).

355   BoM was absent for the metamers (p > 0.05, Wilcoxon signed-rank test). To investigate if the level

356   of BoM differed between the metamers and the original images, we performed a repeated-

357   measures two-way ANOVA with object-border and scrambling (2 levels each) as factors. The main

358   effects of object-borders and scrambling were both significant (salience, $F_{1,97}$ = 28.6, p < 0.001;

359   scrambling, $F_{1,97}$ = 5.42, p = 0.022). Importantly, the interaction was also significant at the population

360   level ($F_{1,97}$ = 6.74, p = 0.011) and at many of the individual recording sites (at p < 0.05; 40% of the

361   sites in monkey B and 73% in monkey M). Hence, if RF stimulus is kept constant, contextual

362   information enhances the V1 activity elicited by object borders, at a latency of ~80ms. These results,

363   taken together, indicate that there are two processes that jointly explain the enhanced activity

364   elicited by object boundaries. The tuning of V1 neurons enhances their representation from an early

365   time point onwards and the scene context causes an additional activity increase at a longer latency.

366

Figure 6. Contextual BoM in V1.
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 panel) or removed the object from the scene by creating *metamers* (middle and right panel). The RF stimulus was kept constant across all the conditions. **b**, Average V1 response elicited by image regions that demarcated object boundaries (red) or were part of the background (blue). Lower panel, response difference. BoM in this condition had a latency of 81ms. **c**, The responses elicited by the metamers revealed no significant differences.

Discussion

We investigated how objects in natural images influence neuronal activity in early visual cortex and

observed widespread influences of objecthood on neuronal activity in the human early visual cortex.

These results were mirrored by the early and late modulation of neuronal activity in areas V1 and

21

381 V4 of monkeys. Early influences were related to the tuning of the neurons, causing object

382 boundaries to elicit more activity than background elements. However, if we held the image

383 elements in the RF constant, image elements that were part of an object also elicited more activity

384 than elements that were part of the background. This contextual influence manifested during a later

385 phase of the neuronal response, which suggests the involvement of feedback from higher areas

386 and/or horizontal interactions within visual areas. Whereas previous studies on figure-ground

387 segregation and contour integration in early visual cortex used well-controlled, but artificial stimuli,

388 the present results demonstrate that these findings generalize to natural vision. The results are in

389 accordance with theories proposing that image elements of figures are labeled by enhanced

390 neuronal activity in early visual cortex to segregate them from the background[10,13,31].

391 Despite the different tasks and recording modalities between humans and monkeys, the neuronal

392 responses in V1 were strikingly similar between the two species (Fig. S2). Our fMRI experiment

393 revealed that BoM for natural images is present in V1 and other areas of early visual cortex. Object

394 regions evoked stronger response than backgrounds in areas V1 and V2 but OBM did not reach

395 significance in a number of higher areas, including hV4. In contrast, in the electrophysiological

396 experiments in monkeys, OBM was present in V1 but even stronger in V4. This discrepancy may be

397 related to differences between species, experimental setups and differences in spiking versus fMRI

398 measures of neural activity[22,32]. Another relevant difference is the larger size of fMRI pRFs in hV4

399 compared to neurophysiologically determined V4 RFs. The larger pRF sizes in hV4 may include more

400 neurons with RFs not on the boundary and thereby dilute the BoM signal. Human fMRI allowed us

401 to link the neural responses to human perception, and the monkey neurophysiological experiments

402 allowed us to measure the timing of BoM and OBM and relate it to previous neurophysiological

403 work with synthetic stimuli.

404

Early and later object boundary signals

406 Unexpectedly, natural images elicited BoM during the initial V1 response, at a latency of 50ms. This

407 is much earlier than in previous studies that used well controlled, but artificial stimuli to keep the

408 RF stimulus identical between salient and non-salient contour conditions[6,7]. In these previous

409 studies, the contextual effects on neuronal firing rates were attributed to feedback from higher

410 cortical areas and/or lateral connections within V1, which can inform neurons about information

411 outside the RF. The synaptic and propagation delays associated with these recurrent routes explain

412 why BoM occurs a few tens of ms after the initial V1 response[13]. Our results indicate that the early

413 BoM signals evoked by natural images are not contextual but reflect the tuning of V1 neurons.

414 Indeed, we found that features of object borders differ from those of non-border image regions (Fig.

415 5c) and that V1 neurons are sensitive to these feature differences (Fig. 5d). On average, the object

416 borders of a particular contrast elicit more activity than non-border image regions with the same

417 contrast. The V1 tuning to object borders is more complex than can be described by Gabor

418 filters[25,28,33] and is presumably related to a sensitivity to higher-order image statistics[34–36], which

419 also explain the early detection of boundaries in studies using synthetic figure-ground displays[11]

420 (Fig. 1b).

421 In addition to their effect on the feedforward response, object boundaries also elicited a contextual

422 influence on V1 activity. When we matched the image elements of object and non-object contours

423 in the RF of V1 neurons, the activity elicited by the object contours was still stronger than that

424 elicited by other, non-object contours (Fig. 6). BoM now occurred at a latency of 81ms, which is

425 30ms later than the feedforward response and in line with previous studies that used synthetic

426 stimuli to keep the RF content constant and controlled contour salience by the layout of image

23

427   elements in the RF surround[6,7]. This additional delay suggests that BoM now depended on feedback

428   from higher areas and/or horizontal connections within V1. It is of interest that these putative

429   feedback signals increased the activity elicited by contours that are predicted by an object's overall

430   shape. This result is not in accordance with popular "predictive coding" schemes[37], which suggest

431   that feedback connections should suppress the activity of contours that are predicted by the

432   object's shape. Instead, we found that object borders increase the neuronal activity in the visual

433   cortex, both during the early and later phases of V1 response.

434   BoM is presumably related to border-ownership coding, which is expressed by many neurons in V2,

435   V3, V4 and also by some V1 neurons[38–40]. The activity of neurons with border-ownership signals

436   depends on the side of the figural region relative to the border that falls in the RF. For example, if

437   the border is vertical, some neurons prefer that the border is owned by a figure on the left of it,

438   whereas other neurons have the opposite preference. Hence, border-ownership neurons can link

439   the shape of the border to the surface properties of the object's interior and may therefore play an

440   important role in object recognition. In many situations, the local shape of a border falling in a RF

441   can provide information about the side of the figure[24]. In these situations, neurons express border-

442   ownership early, during the feedforward response. However, if the RF-stimulus is held constant,

443   border-ownership coding occurs after an additional delay[40]. Although BoM reflects extra activity

444   elicited by the object boundaries compared to less-relevant image elements, and thereby differs

445   from border-ownership coding, it seems likely that the two effects are intimately related.

446

447   **Neuronal activity elicited by object interiors**

448   Image elements that were part of the interior of objects in the scene elicited more activity than

449   background elements, both in human fMRI and monkey neurophysiology. This finding generalizes

450   previous results on the neuronal mechanism of figure-ground perception to natural images (Fig.

451   1b,c)[10–12,16,41–46]. Studies using synthetic stimuli revealed a number of successive processing phases

452   for the processing of texture defined figure-ground stimuli (Fig. 1b, reviewed in ref.[31]). In V1, the

453   first phase is the arrival of the input from the LGN at a latency of ~40ms. This is followed at a latency

454   of ~60ms by boundary enhancement. Boundaries between figure and ground now elicit extra

455   activity in V1, an effect that starts in the superficial layers of cortex[41]. The change in feature values

456   at a boundary between figure and ground can be detected locally (e.g. there is an abrupt change in

457   the texture as in Fig. 1b) and the mechanisms presumably overlap with early BoM (Fig. 7a). In a yet

458   later phase, at a latency of ~90ms, V1 neurons that represent the figure's interior enhance their

459   activity. Enhancement in the figure's interior is a genuine contextual effect, because the properties

460   of the image elements that fall into the RF are often not informative about whether they belong to

461   figure and ground (Fig. 1b,c). In these cases, the information that a RF falls on a figure comes from

462   outside the RF. The relatively long latency of this figure-ground modulation is compatible with

463   recurrent loops that may include horizontal connections within V1 and loops through the higher

464   visual areas. Indeed, if activity in higher areas is blocked, figure-ground modulation in the center of

465   the figure is diminished[47,48], implying an import contribution of recurrent routes through higher

466   visual cortical areas[49]. Interestingly, the optogenetic blockade of the late V1 activity phase with

467   figure-ground modulation selectively impairs figure-ground perception, whereas contrast detection

468   is unimpaired[48].

469   The activity of image elements that were part of the interior of objects of natural images was

470   enhanced in V1 at a latency of 78ms, which is 28ms after the visually driven response during the

471 first fixation. In V4 this OBM signal occurred at a latency of 93ms. We also observed systematic

472 differences between the features of object interiors and those in the background, indicating that

473 the tuning of V1 neurons could, in principle, discriminate between features of figure and

474 background, even though the early V1 population response did not exhibit OBM.

475

476 **Trans-saccadic integration**

477 Previous studies demonstrated that figure-ground signals for synthetic stimuli can persist across eye

478 movements[17,18]. As a result, the figure-ground structure that is perceived during one fixation can be

479 quickly reassigned to the appropriate neurons after all RFs shifted across the image due to the

480 saccade. In the present study, OBM occurred sooner after the visually driven response for later

481 fixations than for the first fixation. This result suggests that information about the location of object

482 interiors is indeed carried over to the new fixation[18], providing insight into the neuronal mechanisms

483 for trans-saccadic integration in natural images[17,50,51]. One possible mechanism for the remapping

484 of these response modulations in early visual cortex are neurons in parietal and frontal cortex that

485 remap salient image elements and could provide feedback to lower areas after each saccade[52–54].

486 Another possible mechanism is provided by neurons that code the position of objects in non-

487 retinotopic, e.g. head-centered coordinates[55,56]. These cells do not need to update their activity

488 after an eye movement because object position relative to the head is independent of eye position.

489 These neurons could feed the location of objects back to early visual cortex after a coordinate

490 transformation from head to eye centered coordinates, based on the new, post-saccadic eye

491 position. A final source for the early post-saccade OBM are neurons in areas of the temporal stream

492 that code for the overall shape of objects. Many of these neurons are translation invariant, i.e. their

493 activity depends little on the precise location of an object on the retina[57]. These neurons represent

494   object shape[58] and could provide feedback to boost the activity of neurons in lower areas that

495   represent relevant shape features after a saccade.

496
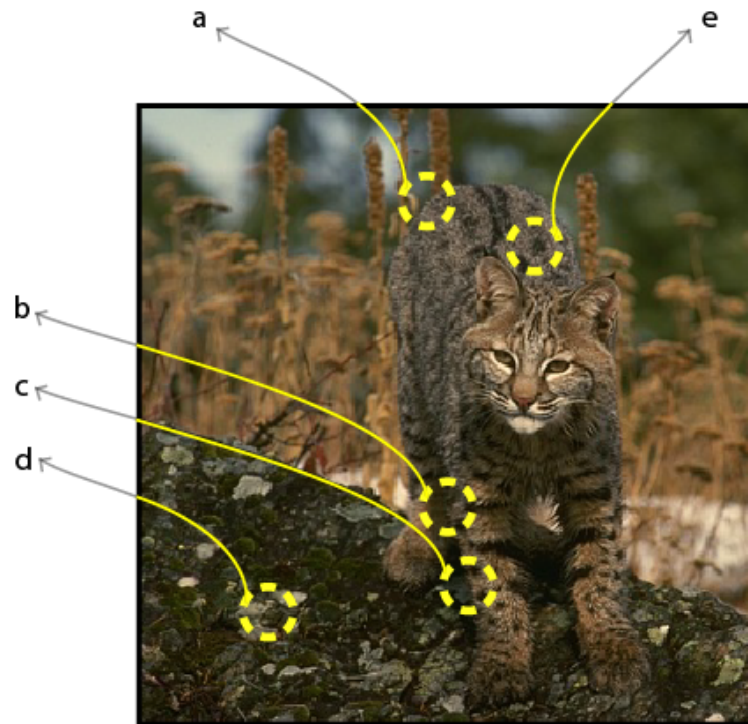
497

498   **Figure 7. Bottom-up and top-down mechanisms for object detection in natural vision**
499   Summary of the results. The local features of contour **a** suggest that is an object boundary. It can be
500   detected bottom-up by tuning of V1 neurons. Image patches **b** and **c** have similar features but the
501   context indicates that **c** contains an object boundary and **b** does not. Image patches **d** and **e** have
502   similar features, but only **e** is part of the interior of the animal.
503

504   **Conclusion**

505   We conclude that the object boundaries and object interiors of natural images increase neuronal

506   activity in the visual cortex. The extra neuronal activity occurs early if the local image elements in

507   the RF have a high degree of "objecthood" and at a later point in time if it depends on contextual

508   information outside the RF. OBM and BoM play an important role in perceptual organization[13], the

509   process that groups image elements of the same object together and segregates them from other

27

510   objects and the background by labeling the object features with enhanced neuronal activity[13,59–63].

511   The presence of these cortical image parsing signals for natural images suggest that they play a role

512   during each fixation of our everyday vision, opening many avenues for future research.

513

520

521   **References**

522   1.   Martin, D., Fowlkes, C., Tal, D. & Malik, J. A database of human segmented natural images

523   and its application to evaluating segmentation algorithms and measuring ecological

524   statistics. in *Proceedings of the IEEE International Conference on Computer Vision* vol. 2

525   416–423 (Iccv Vancouver:, 2001).

526   2.   Zuiderbaan, W., van Leeuwen, J. & Dumoulin, S. O. Change Blindness Is Influenced by Both

527   Contrast Energy and Subjective Importance within Local Regions of the Image. *Front Psychol*

528   **8**, 1718 (2017).

529   3.   Neri, P. Object segmentation controls image reconstruction from natural scenes. *PLoS Biol*

530   **15**, e1002611 (2017).

531   4.   Self, M. W., Mookhoek, A., Tjalma, N. & Roelfsema, P. R. Contextual effects on perceived

532   contrast: figure-ground assignment and orientation contrast. *Journal of Vision* **15**, 2 (2015).

28

533   5.   Papale, P. *et al.* Foreground-background segmentation revealed during natural image

534   viewing. *eNeuro* **5**, (2018).

535   6.   Li, W., Piech, V. & Gilbert, C. D. Contour saliency in primary visual cortex. *Neuron* **50**, 951–

536   962 (2006).

537   7.   Chen, M. *et al.* Incremental integration of global contours through interplay between visual

538   cortical areas. *Neuron* **82**, 682–694 (2014).

539   8.   Sigman, M., Cecchi, G. A., Gilbert, C. D. & Magnasco, M. O. On a common circle: natural

540   scenes and Gestalt rules. *Proc Natl Acad Sci U S A* **98**, 1935–1940 (2001).

541   9.   Geisler, W. S., Perry, J. S., Super, B. J. & Gallogly, D. P. Edge co-occurrence in natural images

542   predicts contour grouping performance. *Vision Res* **41**, 711–724 (2001).

543   10.   Lamme, V. A. The neurophysiology of figure-ground segregation in primary visual cortex. *J*

544   *Neurosci* **15**, 1605–1615 (1995).

545   11.   Poort, J. *et al.* The role of attention in figure-ground segregation in areas V1 and V4 of the

546   visual cortex. *Neuron* **75**, 143–156 (2012).

547   12.   Poort, J., Self, M. W., van Vugt, B., Malkki, H. & Roelfsema, P. R. Texture Segregation Causes

548   Early Figure Enhancement and Later Ground Suppression in Areas V1 and V4 of Visual

549   Cortex. *Cereb Cortex* **26**, 3964–3976 (2016).

550   13.   Roelfsema, P. R. Cortical algorithms for perceptual grouping. *Annu Rev Neurosci* **29**, 203–

551   227 (2006).

552   14.   Roelfsema, P. R. & Houtkamp, R. Incremental grouping of image elements in vision. *Atten*

553   *Percept Psychophys* **73**, 2542–2572 (2011).

554   15.   Felsen, G. & Dan, Y. A natural approach to studying vision. *Nature Neuroscience* **8**, 1643–

555   1646 (2005).

556   16.   Self, M. W. *et al.* The Segmentation of Proto-Objects in the Monkey Primary Visual Cortex.

557   *Curr Biol* **29**, 1019-1029 e4 (2019).

558   17.   Khayat, P. S., Spekreijse, H. & Roelfsema, P. R. Correlates of transsaccadic integration in the

559   primary visual cortex of the monkey. *Proc Natl Acad Sci U S A* **101**, 12712–12717 (2004).

560   18.   O'Herron, P. & von der Heydt, R. Remapping of border ownership in the visual cortex. *J*

561   *Neurosci* **33**, 1964–1974 (2013).

562   19.   Dumoulin, S. O. & Wandell, B. A. Population receptive field estimates in human visual

563   cortex. *NeuroImage* **39**, 647–660 (2008).

564   20.   Zuiderbaan, W., Harvey, B. M. & Dumoulin, S. O. Image identification from brain activity

565   using the population receptive field model. *PLoS ONE* **12**, e0183295 (2017).

566   21.   Arbelaez, P., Maire, M., Fowlkes, C. & Malik, J. Contour detection and hierarchical image

567   segmentation. *IEEE Trans Pattern Anal Mach Intell* **33**, 898–916 (2011).

568   22.   Self, M. W., van Kerkoerle, T., Goebel, R. & Roelfsema, P. R. Benchmarking laminar fMRI:

569   Neuronal spiking and synaptic activity during top-down and bottom-up processing in the

570   different layers of cortex. *NeuroImage* **197**, 806–817 (2019).

571   23.   Roelfsema, P. R., Khayat, P. S. & Spekreijse, H. Subtask sequencing in the primary visual

572   cortex. *Proc Natl Acad Sci U S A* **100**, 5467–5472 (2003).

573   24.   Fowlkes, C. C., Martin, D. R. & Malik, J. Local figure-ground cues are valid for natural images.

574   *Journal of Vision* **7**, 2 (2007).

575   25.   Walker, E. Y. *et al.* Inception loops discover what excites neurons most using deep

576   predictive models. *Nature Neuroscience* (2019) doi:10.1038/s41593-019-0517-x.

577   26.   Bashivan, P., Kar, K. & DiCarlo, J. J. Neural population control via deep image synthesis.

578   *Science* (2019) doi:10.1126/science.aav9436.

579   27.   Klindt, D. A., Ecker, A. S., Euler, T. & Bethge, M. Neural system identification for large

580   populations separating "what" and "where." in *Advances in Neural Information Processing*

581   *Systems* (2017). doi:10.12751/nncn.bc2017.0132.

582   28.   Cadena, S. A. *et al.* Deep convolutional models improve predictions of macaque V1

583   responses to natural images. *PLOS Comput Biol* **15**, e1006897 (2019).

584   29.   Zhang, Y., Lee, T. S., Li, M., Liu, F. & Tang, S. Convolutional neural network models of V1

585   responses to complex patterns. *J Comput Neurosci* **46**, 33–54 (2019).

586   30.   Portilla, J. & Simoncelli, E. P. A parametric texture model based on joint statistics of complex

587   wavelet coefficients. *International Journal of Computer Vision* **40**, 49–70 (2000).

588   31.   Roelfsema, P. R. & de Lange, F. P. Early Visual Cortex as a Multiscale Cognitive Blackboard.

589   *Annu Rev Vis Sci* **2**, 131–151 (2016).

590   32.   O'Herron, P. *et al.* Neural correlates of single-vessel haemodynamic responses in vivo.

591   *Nature* **534**, 378–382 (2016).

592   33.   Tang, S. *et al.* Complex Pattern Selectivity in Macaque Primary Visual Cortex Revealed by

593   Large-Scale Two-Photon Imaging. *Curr Biol* **28**, 38-48 e3 (2018).

594   34.   Johnson, A. P. & Baker, C. L. First- and second-order information in natural images: a filter-

595   based approach to image statistics. *Journal of the Optical Society of America A* **21**, 913

596   (2004).

597   35.   Dimattina, C. & Baker, C. L. Modeling second-order boundary perception: A machine

598   learning approach. *PLoS Computational Biology* **15**, e1006829 (2019).

599   36.   Mareschal, I. & Baker, C. L. A cortical locus for the processing of contrast-defined contours.

600   *Nature Neuroscience* **1**, 150–154 (1998).

601 37. Friston, K. A theory of cortical responses. *Philosophical Transactions of the Royal Society B:*

602 *Biological Sciences* **360**, 815–836 (2005).

603 38. Williford, J. R. & von der Heydt, R. Figure-Ground Organization in Visual Cortex for Natural

604 Scenes. *eNeuro* **3**, (2016).

605 39. Zhou, H., Friedman, H. S. & von der Heydt, R. Coding of border ownership in monkey visual

606 cortex. *J Neurosci* **20**, 6594–6611 (2000).

607 40. Hesse, J. K. & Tsao, D. Y. Consistency of Border-Ownership Cells across Artificial Stimuli,

608 Natural Stimuli, and Stimuli with Ambiguous Contours. *J Neurosci* **36**, 11338–11349 (2016).

609 41. Self, M. W., van Kerkoerle, T., Super, H. & Roelfsema, P. R. Distinct roles of the cortical

610 layers of area V1 in figure-ground segregation. *Curr Biol* **23**, 2121–2129 (2013).

611 42. Scholte, H. S., Jolij, J., Fahrenfort, J. J. & Lamme, V. A. Feedforward and recurrent processing

612 in scene segmentation: electroencephalography and functional magnetic resonance

613 imaging. *J Cogn Neurosci* **20**, 2097–2109 (2008).

614 43. Super, H., Spekreijse, H. & Lamme, V. A. Two distinct modes of sensory processing observed

615 in monkey primary visual cortex (V1). *Nature Neuroscience* **4**, 304–310 (2001).

616 44. Zipser, K., Lamme, V. A. & Schiller, P. H. Contextual modulation in primary visual cortex. *J*

617 *Neurosci* **16**, 7376–7389 (1996).

618 45. Likova, L. T. & Tyler, C. W. Occipital network for figure/ground organization. *Exp Brain Res*

619 **189**, 257–267 (2008).

620 46. Poltoratski, S. & Tong, F. Resolving the spatial profile of figure enhancement in human V1

621 through population receptive field modeling. *Journal of Neuroscience* (2020)

622 doi:10.1523/JNEUROSCI.2377-19.2020.

623 47. Klink, P. C., Dagnino, B., Gariel-Mathis, M. A. & Roelfsema, P. R. Distinct Feedforward and

624 Feedback Effects of Microstimulation in Visual Cortex Reveal Neural Mechanisms of Texture

625 Segregation. *Neuron* **95**, 209-220 e3 (2017).

626 48. Kirchberger, L. *et al.* The essential role of recurrent processing for figure-ground perception

627 in mice. *Science Advances* **7**, (2021).

628 49. Kar, K., Kubilius, J., Schmidt, K., Issa, E. B. & DiCarlo, J. J. Evidence that recurrent circuits are

629 critical to the ventral stream's execution of core object recognition behavior. *Nature*

630 *Neuroscience* **22**, 974–983 (2019).

631 50. McConkie, G. W. & Currie, C. B. Visual stability across saccades while viewing complex

632 pictures. *J Exp Psychol Hum Percept Perform* **22**, 563–581 (1996).

633 51. Irwin D.J., D. E. ; G. Eye movements, attention and trans-saccadic memory. *J Visual*

634 *Cognition* **5**, 127–155 (1998).

635 52. Gottlieb, J. P., Kusunoki, M. & Goldberg, M. E. The representation of visual salience in

636 monkey parietal cortex. *Nature* **391**, 481–484 (1998).

637 53. Duhamel, J. R., Colby, C. L. & Goldberg, M. E. The updating of the representation of visual

638 space in parietal cortex by intended eye movements. *Science* **255**, 90–92 (1992).

639 54. Umeno, M. M. & Goldberg, M. E. Spatial processing in the monkey frontal eye field. I.

640 Predictive visual responses. *J Neurophysiol* **78**, 1373–1383 (1997).

641 55. Duhamel, J. R., Bremmer, F., ben Hamed, S. & Graf, W. Spatial invariance of visual receptive

642 fields in parietal cortex neurons. *Nature* **389**, 845–848 (1997).

643 56. Graziano, M. S., Yap, G. S. & Gross, C. G. Coding of visual space by premotor neurons.

644 *Science* **266**, 1054–1057 (1994).

645 57. Tovee, M. J., Rolls, E. T. & Azzopardi, P. Translation invariance in the responses to faces of

646 single neurons in the temporal visual cortical areas of the alert macaque. *J Neurophysiol* **72**,

647 1049–1060 (1994).

648 58. Bao, P., She, L., McGill, M. & Tsao, D. Y. A map of object space in primate inferotemporal

649 cortex. *Nature* **583**, 103–108 (2020).

650 59. Lamme, V. A. & Roelfsema, P. R. The distinct modes of vision offered by feedforward and

651 recurrent processing. *Trends Neurosci* **23**, 571–579 (2000).

652 60. Houtkamp, R. & Roelfsema, P. R. Parallel and serial grouping of image elements in visual

653 perception. *J Exp Psychol Hum Percept Perform* **36**, 1443–1459 (2010).

654 61. Pooresmaeili, A. & Roelfsema, P. R. A growth-cone model for the spread of object-based

655 attention during contour grouping. *Curr Biol* **24**, 2869–2877 (2014).

656 62. Korjoukov, I. *et al.* The time course of perceptual grouping in natural scenes. *Psychol Sci* **23**,

657 1482–1489 (2012).

658 63. Jeurissen, D., Self, M. W. & Roelfsema, P. R. Serial grouping of 2D-image regions with

659 object-based attention in humans. *Elife* **5**, (2016).

660

661

34

Methods

fMRI experiment with human participants

Subjects

Four participants (all male; ages 29-41 years) 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[64,65] on a Macintosh MacBook Pro. The stimuli were back-projected on a display inside the MRI bore. The subject viewed the display through mirrors inside the scanner. The size of the display was 15.0x7.9cm with a resolution of 1024x538 pixels. The total distance from the subject's eyes to the display was 41cm. 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.

Population receptive field (pRF) mapping stimulus

We used bar apertures filled with natural images [19,20] (Fig. S2) to train the pRF-model. 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

685 duration of 1.5 seconds per bar position) so that each pass took 30 seconds. A period of 30 seconds

686 mean-luminance (0% contrast) was presented after every pass. In total there were 4 blocks of mean-

687 luminance during each scan, presented at evenly spaced intervals. The participants performed a

688 fixation dot task to make sure they fixated at the center of the display. A small fixation dot (0.11°

689 radius) was presented in the middle of the stimulus. The fixation dot changed its color from red to

690 green at random time intervals and subjects were instructed to respond to color changes using a

691 button press.

692

693 *Natural images*

694 The natural images came from the BSD [1,21]. The original resolution of the images was 321x481 pixels

695 (both landscape and portrait). In the fMRI experiments[20], we selected a square region of 321x321

696 pixels from the images and upsampled it to a resolution of 516x516 pixels, which corresponds to a

697 stimulus of 11x11° diameter of visual angle. The images were masked by a circle with a raised cosine

698 faded edge (width of 0.9°), and the areas outside this circle were set to the mean luminance. The

699 images were gamma-linearized and the mean contrast was set to 50%. We used 3 image sets in

700 different scanning runs, each containing 15 different natural images (45 in total) and one full-field

701 binarized bandpass-filtered noise stimulus. Figure S1 shows the image set. A fixation dot was

702 presented at the center of the stimulus. We used the same fixation dot task as for the pRF mapping

703 runs.

704

705 *Functional imaging and processing*

706 The MRI data was acquired with a Philips 7T scanner using a 32-channel head-coil[20]. We scanned

707 the participants with a 2d-echo-planar-imaging sequence with 25 slices oriented perpendicular to

36

708   the calcarine sulcus with no gap. The following parameters were used; repetition time (TR) =

709   1500ms, echo time (TE) = 25ms and a flip angle of 80°. The functional resolution was 2x2x2mm and

710   the field of view (FOV) was 190x190x50mm. We used foam padding to minimize head movement.

711   The functional images were corrected for head movement between and within the scans[66]. For

712   computation of the head movement between scans, the first functional volumes for each scan were

713   aligned. Within scan motion correction was then computed by aligning the frames of a scan to the

714   first frame. The duration of the pRF mapping scans was 372 seconds (248 time-frames), of which

715   the first 12 seconds (8 time-frames) were discarded to eliminate start-up magnetization transients.

716   During the three sessions we acquired 6-8 pRF mapping scans in total per subject. To obtain a high

717   signal-to-noise ratio, we averaged across the repeated scans. During the three sessions in which we

718   presented the natural images we acquired 6-7 scans for each of the three stimulus sets. The duration

719   of the scans with the natural images was 432 seconds (288 time-frames). The first 12 seconds (8

720   time-frames) were discarded to eliminate start-up magnetization transients. The images were

721   presented in a block design. Each image was presented during a 9-second block. Within this block

722   the same image was shown 18 times for a duration of 300ms followed by 200ms mean-luminance.

723   The full-field stimuli were presented with 3 alternating different high-contrast patterns, to obtain a

724   full high-contrast response that is not based upon one specific high-contrast pattern (Fig. S2b).

725   Specifically, the phase of the full-field pattern was randomized on different presentations in order

726   to obtain a response that is not influenced by one specific dartboard pattern. The block in which the

727   stimulus was presented was followed by a 12 second mean-luminance presentation. Four longer

728   blank periods of 33 seconds were also included during the scan.

729

730   *Anatomical imaging and processing*

731   The T1-weighted MRI images were acquired in a separate session using an 8-channel SENSE head-

732   coil. The following parameters were used: TR/TE/flip angle = 9.88/4.59/8. The scans were acquired

733   at a resolution of 0.79x0.80x0.80mm and were resampled to a resolution of 1mm^3 isotropic. The

734   functional MRI scans were aligned with the anatomical MRI using an automatic alignment

735   technique[66]. From the anatomical MRI, white matter was automatically segmented using the

736   FMRIB's Software Library (FSL)[67]. After the automatic segmentation it was hand-edited to minimize

737   segmentation errors[68]. The gray matter was grown from the white matter to form a 4mm layer

738   surrounding the white matter. A smoothed 3D cortical surface can be rendered by reconstruction

739   of the cortical surface at the border of the white and gray matter[69].

740

741   *pRF model-based analysis*

742   The pRF-model was estimated for every cortical location from the measured fMRI signal that was

743   elicited by the pRF mapping bar stimuli (Fig. S2a)[19,20]. In short, the method estimates the pRF by

744   combining the measured fMRI time-series with the position time course of the visual stimulus. A

745   prediction of the time-series is made by calculating the overlap of the pRF and the stimulus energy

746   (RMS contrast, see below) convolved with the hemodynamic response function (HRF). We

747   estimated the parameters of the HRF that best describes the data of the whole acquired fMRI

748   volume[70]. The optimal parameters of the pRF-model are chosen by minimizing the residual sum of

749   squares between the predicted and the measured time-series. We used the conventional pRF-

750   model, which consists of a circular symmetric Gaussian. This model has four parameters: position

751   (x, y), size (σ) and amplitude (β). For further technical and implementation details see[19].

752

753   *Regions of interest*

754   We used the pRF-method to estimate position parameters x, and y of the pRF of every voxel. From

755   these values, we derived the polar angle ($atan(y_0/x_0)$) and eccentricity ($\sqrt{(x_0^2 + y_0^2)}$) values. We drew

756   the borders between visual field maps on the basis[71] polar angle and eccentricity maps on the

757   inflated cortical surface[69]. We defined visual areas V1, V2, V3, hV4, LO-1/2 and V3-a/b as our regions

758   of interest (ROIs)[71–74].

759

760   *Analysis of fMRI responses to the natural images*

761   We measured fMRI responses to 45 natural images (Fig. S1) and 3 full-field high contrast stimuli

762   (100% contrast; Figure S2)[20]. We first determined the voxel response amplitudes in %BOLD signal

763   change elicited by each of these images. The voxel responses were calculated using a general linear

764   model (GLM)[75,76]. To reduce the noise from the individual voxel differences in response amplitudes,

765   we normalized the responses to the voxel's response to the full-field (100% contrast) stimulus.

766   To determine the contrast response function (CRF), we only used the voxels with an overall

767   significant response (t-values > 4.0), a pRF eccentricity between 0.5 and 4° and for which the pRF

768   model explained more than 40% of the variance. Based on previous work, for every area we used a

769   threshold for the pRF sizes[19,70,77,78]. In V1 we included pRFs with a value of σ (which determines pRF

770   size) between 0.25° and 0.8°, for V2 between 0.25° and 1.1°, for V3 between 0.25° and 1.75°, for

771   hV4 between 0.45° and 3°, for V3ab between 0.45° and 3.75° and for and LO12 between 0.9° and

772   5°.

773   To derive the CRF, we computed the contrast of every natural image within each pRF. The pRF of

774   voxels was modeled as a circular symmetric Gaussian function, described by parameters for position

775   (x_c, y_c) and size (σ), giving rise to a Gaussian weighting function w_i:

776
$$w_i = exp\left(\frac{(x_i - x_c)^2 + (y_i - y_c)^2}{2\sigma^2}\right) \qquad (1)$$

777 Where x_c and y_c define the location of the center of the pRF in the visual field, σ determines the size

778 of the pRF and x_i and y_i define the location of the i-th pixel. We computed each voxel's contrast

779 value to each natural image by calculating the Root-Mean-Squared (RMS) contrast[65,79] of the part

780 of the image inside the voxel's pRF. RMS contrast was defined as the standard deviation of the

781 luminance of the pixels relative to the mean. The RMS-contrast was weighted by the pRF Gaussian

782 function to obtain the local contrast-energy value per pRF:

783
$$local\ contrast\ energy = \sqrt{\frac{1}{\sum_{i=1}^{N} w_i} \sum_{i=1}^{N} w_i \frac{(L_i - L)^2}{L^2}} \qquad (2)$$

784 Where N is the number of pixels in the stimulus window. L is the mean luminance from the pixels

785 inside the spatial window, and L_i is the luminance of the i-th pixel[2].

786 We computed the CRF of voxels areas V1, V2, V3, hV4, LO-1/2 and V3-a/b by measuring the fMRI

787 responses as a function of the contrast inside the pRF. We chose contrast bins such that every bin

788 contained 10% of the voxels and fitted the following equation (modified from ref.[80]):

789
$$R(C) = a\frac{C^q}{C^q + Q^q} \qquad (3)$$

790 where R is the fMRI response, C is the RMS-contrast inside the pRF, Q represents the contrast value

791 where the CRF is at half of its maximum response, and q determines the slope (Q and q are free

792 parameters).

793

794 *Quantification of BoM and OBM*

795 The BSD images are annotated by 5-9 human observers who drew lines to identify borders that are

796 important for the scene's representation[1,21]. We used these measurements to define the perceived

40

797 boundaries, which are salient boundaries of the scene. Every pixel i of the manually labeled images

798 have values for the degree of agreement between observers, S_i, between 0 (not labeled by any

799 observer) and 1 (labeled by all observers). The border-salience in the pRF is calculated as a weighted

800 sum across pixels:

801
$$Perceived\ border = \frac{\sum_{i=1}^{N} w_i \cdot S_i}{\sum_{i=1}^{N} w_i} \qquad (4)$$

802 Here w_i are the weights of the RF estimate (equation 1) and N is the total number of pixels in the

803 RF. We used the same method to quantify the degree to which a pixel was part of an object or the

804 background (Fig. 1e). Pixels that were as part of an object, had a value of 1 and pixels that were part

805 of the background had a value of 0. We selected a segmentation covering the objects of a scene

806 from one of the BSD subjects, and then considered everything else as background[5]. We excluded 3

807 of the 45 images in the OBM analysis because the object in the image almost filled the entire scene.

808 We split the voxels based on objecthood values inside the pRFs. We included the lowest 25 percent

809 responses as non-perceived borders/background and the highest 25 percent as perceived-

810 borders/object-interior and computed the CRFs within these voxel classes.

811

812 *Statistics*

813 We used a bootstrapping procedure to determine the significance of differences in CRFs between

814 conditions. We sampled the images with replacement 1000 times, fit the CRF for the two simulated

815 conditions and computed the mean difference. We derived the p-value from this null distribution.

816

817 **Electrophysiological experiments in monkeys**

818 *Training of the monkeys*

41

819   All procedures complied with the NIH Guide for Care and Use of Laboratory Animals and were

820   approved by the institutional animal care and use committee of the Royal Netherlands Academy of

821   Arts and Sciences. Two macaque monkeys (males, 7 and 13 years old) participated in the

822   electrophysiological experiments. They were socially housed in stable pairs in a specialized primate

823   facility with natural daylight, controlled humidity and temperature. The home-cage was a large

824   floor-to-ceiling cage which allowed natural climbing and swinging behavior. The cage had a solid

825   floor, covered with sawdust and was enriched with toys and foraging items. Their diet consisted of

826   monkey chow supplemented with fresh fruit. Their access to fluid was controlled, according to a

827   carefully designed regime for fluid uptake. During weekdays the animals received water or diluted

828   fruit juice in the experimental set-up upon correctly performed trials. We ensured that the animals

829   drank sufficient fluid in the set-up and supplemented extra fluid after the recording session if they

830   did not drink enough. On days of the weekend, they received at least 700ml water in the home-cage

831   in a drinking bottle. The animals were regularly checked by veterinary staff and animal caretakers

832   and their weight and general appearance were recorded daily in an electronic logbook during fluid-

833   control periods.

834

835   *Surgical details*

836   We implanted both monkeys with a titanium head-post (Crist instruments) under aseptic conditions

837   and general anesthesia as reported previously[81]. The monkeys were trained to direct their gaze to a

838   0.5° diameter fixation dot and hold their eyes within a fixation window (1.1° diameter). They then

839   underwent a second operation to implant 5x5 arrays of micro-electrodes (Utah-probes, Blackrock

840   Microsystems) over opercular V1 and V4. The inter-electrode spacing of the arrays was 400μm. We

841   obtained good signals from 4 V1 arrays in each monkey and from 2 V4 arrays in monkey B[11].

42

842

843 *Electrophysiology*

844 We recorded neuronal activity of 192 recording sites in V1 (96 in Monkey M and 96 in Monkey B)

845 and 48 V4 recording sites in monkey B. We recorded the envelope of multi-unit activity by digitizing

846 the signal referenced to a subdural electrode at 24.4kHz. The signal was band-pass filtered (2nd

847 order Butterworth filter, 500Hz-5KHz) to isolate high-frequency (spiking) activity. This signal was

848 rectified (negative becomes positive) and low-pass filtered (corner frequency = 200Hz) to produce

849 the envelope of the high-frequency activity, which we refer to as MUA[82]. The MUA signal reflects

850 the population spiking of neurons within 100-150μm of the electrode and the population responses

851 are very similar to those obtained by pooling across single units[82–85].

852

853 *Receptive Field Mapping*

854 We mapped the RF of each recording site in V1 using a drifting luminance-defined bar that moved

855 in one of four directions. The response to each direction was fitted with a Gaussian function. The

856 borders of the RF were then calculated as described previously[82]. The signal-to-noise ratio (SNR_{RF})

857 of the response was taken as the peak of the Gaussian divided by the standard deviation of the pre-

858 trial baseline response. We only included recording sites in the analyses with a reliable visual

859 response (i.e., the responses to all four bar directions had an SNR_{RF} of at least 1). The median V1 RF

860 size, taken as the square-root of the area, was 1.8° (range 0.4° to 8.2°) and the median eccentricity

861 of the RFs was 2.4° (range 0.6° to 12.9°). We mapped V4 RFs by presenting white dots (0.5°,

862 luminance 82 cd·m^{-2}) on a gray background (luminance 14 cd·m^{-2}) at different positions of a grid

863 (0.5° spacing). The hotspot of the V4 RF was defined as the position with the maximum response

864 (median eccentricity 4.04°, range 0.79°–7.43°) and the RF borders as the locations where activity

43

865   fell below 50% of the maximum[86]. Using this criterion, the median V4 RF size was 4.5° (range 2.6°–

866   6.0°).

867

868   *Stimulus presentation*

869   In the experiments with monkeys, stimuli were presented on a CRT monitor at a refresh rate of 60Hz

870   and resolution of 1024x768 pixels viewed from a distance of 46cm. The monitor had a width of

871   40cm, yielding a field-of-view of 41.6 x 31.2°. All stimuli were generated in Matlab using the COGENT

872   graphics toolbox (developed by John Romaya at the LON at the Wellcome Department of Imaging

873   Neuroscience). The eye position was recorded using a digital camera (Thomas recordings, 250Hz

874   frame-rate).

875

876   *Selection of recording sites and inclusion of data*

877   To normalize MUA, we first subtracted the mean activity in the pre-trial period in which the animal

878   was fixating (200 to 0ms relative to stimulus onset) and divided by the maximum smoothed (26ms

879   Gaussian kernel) peak response (0-150ms after stimulus onset). In the experiment with multiple

880   saccades, each trial contained multiple fixations and neuronal activity was normalized to the peak

881   response elicited by stimulus onset during the first fixation. The data are therefore in normalized

882   units, where e.g. a value of 0.1 indicates 10% of the maximal MUA onset response. We only included

883   recording sites on days with a sufficient signal-to-noise ratio (SNR_{DAY}). SNR_{DAY} was estimated by

884   dividing the maximum of the initial peak response by the standard deviation of the baseline activity

885   across trials. When the SNR_{DAY} of a recording site was smaller than 2 on particular day, we removed

886   that session from the analysis of that recording site. To test for statistical differences between

887 conditions and to compute the CRFs, MUA activity was generally averaged in a 0-300ms time

888 window. Analyses with different time-windows have been specified in the main text.

889

890 *Analyses of latency*

891 To compute the latency of neural responses a function was fitted to the time-course of interest (i.e.

892 the difference between object borders and non-border image regions or the difference between

893 the object interior and background)[11,12,23]. The function was derived from the assumptions that the

894 onset of the response has a Gaussian distribution and that a fraction of the response dissipates

895 exponentially which yields the following equation:

896 $$f(t) = d \cdot \exp\left(\mu\alpha + 0.5\sigma^2\alpha^2 - \alpha t\right) \cdot \left(G(t, \mu + \sigma^2\alpha, \sigma) + c \cdot G(t, \mu, \sigma)\right) \tag{5}$$

897 Where $G(t, \mu, \sigma)$ is a cumulative Gaussian density with mean μ and standard deviation σ, α^{-1} is the

898 time constant of the dissipation, and c and d represent the contribution the non-dissipating and

899 dissipating components, respectively. The latency was defined as the point at which the fitted

900 function reached 33% of its maximum. To compare the latency of the BoM and OBM between

901 fixation 1 and fixations 2-6, we first subtracted from the OBM latency for each recording site from

902 the latency of visually driven response and performed a Wilcoxon signed rank test (Fig. S7). The

903 latency of visually driven response was computed as the difference between the response elicited

904 by images with the highest and lowest contrast levels in the RF.

905

906 *Natural images presented in the electrophysiological experiments*

907 Four BSD images from the fMRI experiment were used in the electrophysiological experiments in

908 which the monkeys made saccades (11.6° radius visual angle; Figure S1). At the start of the trial the

909 screen was gray (26.8 cd·m^{-2}) with a red fixation point with a position that was randomly selected

910 from uniformly spaced grid (with ~500 positions) covering the circular aperture of the image. The

911 image appeared once the monkey had maintained fixation for 300ms (fixation 1). After an additional

912 400ms, the first fixation point disappeared and another fixation point appeared, at a position

913 sampled from the same grid. The monkey made a saccade to the new fixation point and maintained

914 fixation for an additional 400ms. This fixation-saccade procedure was repeated five times (fixations

915 2-6). Reward was delivered after every correct fixation, with an extra amount at the end of the trial,

916 i.e. after the 6th correct fixation. Aborted trials (i.e., when the monkeys did not maintain fixation for

917 400ms or did not perform a saccade within 700ms) were repeated at the end. The same image was

918 presented in multiple recording days until data for five repetitions of each grid point for every

919 fixation number was collected. We included data from all correct fixations (e.g., if the trial was

920 interrupted after five fixations, we included the first four). Between the trials, the monkeys

921 occasionally also fixated on parts of the image for longer than 300ms, and we also included these

922 spontaneous fixations in the analysis. We collected a total of 11,783 correct trials for monkey M and

923 13,373 for monkey B, for a total of 50,849 fixations analyzed for monkey M and 60,211 for monkey

924 B.

925

926 *Data analysis*

927 We determined the coordinates of the RF on the image for every fixation and analyzed the data

928 from the first fixation and later fixations separately. We computed contrast, BoM and OBM in the

929 RF, as described above. To quantify the independent influence of object borders, object interiors

930 and contrast, we carried out a variance partitioning analysis[87]. For each recording site, we

931 determined how much variance (R^2) was explained by RMS contrast, object borders and object

932 interiors with independent linear regressions, and by combinations of the three predictors in

933 multiple linear regressions. For example, the independent fraction of explained variance (FEV) for

934 the contrast predictor was computed for every recording site as follows:

935
$$R^2_{contrast} = \frac{\left(R^2_{full} - R^2_{BoM + OBM}\right)}{R^2_{full}} \qquad (6)$$

936 where R^2_{full} is the variance explained by the full model, including all the predictors, while

937 $R^2_{BoM + OBM}$ is the variance explained by the model including BoM and OBM as predictors while

938 leaving contrast out. Similar equations were derived for the FEV accounted for by BoM and OBM.

939 The explained variance estimates were then averaged across recording sites. The full model

940 explained 6.4% of the variance in V1 for fixation 1 (mean across recording sites), 4.3% for fixations

941 2-6, 3.2% in V4 during fixation 1 and 2.5% for fixations 2-6. The FEV values for each area and

942 condition presented in the main text were normalized to these values (see equation 6).

943

944 *RF models and the prediction of perceived borders*

945 We determined the selectivity of the neurons at a recording site (time-window 25-75ms), according

946 to previous studies[25–28] which established a mapping between an artificial neural network (ANN)

947 and neuronal tuning (Fig. S9).

948 We extracted the activity of units of VGG-19's layer conv3_1 (state of the art in predicting V1

949 responses to natural images[28,29]) and followed the approach of ref.[28] with two modifications. We

950 used a two-step mapping[26,27], described by following the equation:

951
$$r = f_{VGG-19}(input) * W_s * W_d \qquad (7)$$

952 where r is the predicted response of V1 recording site, $f_{VGG-19}(input)$ is the output of VGG-19's

953 conv3_1 to our stimulus set (i.e. $input$), and W_s and W_d are two sets of weights defining spatial and

47

954 feature selectivity, respectively. The spatial mask (W_s, initialized as the 2D Gaussian RF estimate)

955 approximates the RF and a weighted sum of the nodes in the ANN (W_d) approximates the feature

956 selectivity of the recorded neurons[26]. We trained the model to optimize W_s and W_d to predict V1

957 responses to the training set (i.e., *input* in eq. 7). The activity of V1 recording sites depends on a

958 small and localized portion of the *input* and we therefore cropped the RF models around the most

959 active pixels (3 SD or more away from the global mean) following the procedure in ref.[25]. To visualize

960 tuning (Fig. 5b,c), we kept W_s and W_d constant and varied *input* to maximize the response of the

961 model for a particular V1 recording site. We used cross-validation to assess the quality of the fit as

962 in ref.[25]. Specifically, we used 5,000 trials for training and 100 trials for cross-validation. We trained

963 the model for 300 epochs with a batch size of 256, using 10% of the training set for validation.

964 To estimate how well the V1 RF models could detect perceived borders (F-stat metric; Fig. 5b), we

965 first convolved the RF models with unseen images from the BSD test set (100 grayscale images), and

966 matched them with the annotated versions[21]. For simplicity, we defined the border detection

967 performance (BoP) as the F-measure, employed by the authors of the BSD for benchmark

968 evaluation, which we computed using the MATLAB code associated with the dataset:

969 (https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/code/)[20]. We estimated the

970 chance-level performance in border-detection with a (null) permutation distribution (horizontal

971 dashed line in Fig. 5c; 97.5% level the distribution), shuffling the class labels (after 72 iterations, one

972 for each recording site included in this analysis. A generalized Pareto distribution was fit to the tail

973 of the permutation distribution[88]. The models were implemented using custom Python code using

974 NumPy[89], SciPy (SciPy.org), Tensorflow 1.5[90] and with modules from

975 https://github.com/dicarlolab/npc[26] and https://github.com/sacadena/Cadena2019PlosCB[28].

976

977   *Perceived borders and object interior detection of a population of V1 neurons*

978   To examine the strength of BoM and OBM signals across a larger population of V1 recording sites

979   (Fig. 5d), we trained SVMs to discriminate between object borders and non-border image regions

980   based on the activity of 19 recording sites with RFs smaller than 1.5°. We also trained them to

981   distinguish between object interiors and the background. We used 2 of the 4 images for training and

982   the other two for cross-validation.

983

984   *Neuronal activity profiles across the images*

985   To examine the overall activity level elicited by the images (Fig. 3b), we multiplied the activity by

986   2d-Gaussian approximation of the RFs, weighted by sampling of the visual space caused by the

987   overall pattern of fixations[5,20,91–95] at 7 time points (from 0 to 300ms in 50ms steps) and activity was

988   averaged within a 25ms window centered on each time-point.

989

990   *Statistics*

991   We compared differences between the CRFs between object borders and non-border image regions

992   and between object interiors and the background using a bootstrapping procedure (1,000

993   iterations), as described for the fMRI data above. To test for differences in the median latencies of

994   BoM and OBM between regions and conditions, we used a signed-rank Wilcoxon test across

995   recording sites. The significance of the Pearson's correlation between BoM at the peak of response

996   and the segmentation performance across V1 recordings sites (Fig. 5b) was assessed with a t-test.

997

998   **Isolated patches experiment**

49

999   To test whether isolated image patches from the BSD that were either centered on object borders

1000   or not elicited a different level of V1 activity, we carried out an additional experiment in monkey B

1001   (50 recording sites, Fig. 5e-g). We chose three V1 recording arrays and centered 100 patches of the

1002   image from the BSD that contained object contours and 100 patches that did not contain object

1003   contours on the RFs. These patches were automatically selected so that the RMS contrast was the

1004   same (70±1%) and the size matched to the median RF of the recording sites of the array (0.9° - 2.0°).

1005   The patches were presented on a grey background (26.8 cd·m^{-2}) while the monkey maintained gaze

1006   on a red fixation point for 300ms. We repeated each stimulus five times and collected a total of

1007   3,000 trials (1,000 trials per array). We tested the significance of the difference in the activity elicited

1008   by isolated object and non-object contour patches at the peak of the response (25-75ms) with a

1009   Wilcoxon signed rank test across recording sites.

1010

1011   **Contextual BoM experiment**

1012   To examine differences in activity elicited by object and non-object contours when the stimulus in

1013   the RF was held constant (Figure 6) we selected twelve images from the BSD, which were cropped

1014   and upsampled to 512 x 512 pixels (23.2° x 23.2°). We ensured that the portion of the image covered

1015   by the RF of each recording site and its surround were exactly the same across conditions (same size

1016   and content, Fig. 6), so that border salience only depended on information outside the neurons' RF.

1017   We used a 2x2 design. The first factor was whether the image element in the RF fell on an object

1018   border (Fig. 6a). The second factor was whether we presented the original image or a scrambled

1019   version (also known as *metamer*). To this aim, we created three further stimuli from each image.

1020   First, we copied a circular patch (80 pixels in diameter, 3.7°) from an object contour location onto a

1021   non-object contour location using Adobe Photoshop (blue circle in Fig. 6a, see Fig. S10 for other

50

L022   example images). The border of this circular patch was smoothed to blend it in at the new location.

L023   We created two metamers using the algorithm of ref.[96], with Matlab code provided by the authors

L024   (https://github.com/freeman-lab/metamers). The two metamers were constructed so that either

L025   the object- or non-object contour was kept intact, with a smooth transition to the surround.

L026   Trials started with a red fixation point and the stimulus appeared after 300ms of fixation. The

L027   monkeys maintained fixation for an additional 400ms after stimulus onset (Fig. 6b). We ensured

L028   that the RFs of V1 recording sites were centered on the image patch, which was identical in the four

L029   conditions. The order of the conditions was randomized across trials and aborted trials (when the

L030   monkeys broke fixation) were repeated at the end. We collected a total of 8,094 trials in monkey M

L031   and 9,111 in monkey B.

L032   We tested the significance of the BoM in a window from 0-300ms after stimulus onset (subtracting

L033   spontaneous activity, -100-0ms) with a Wilcoxon signed rank test across recording sites. We also

L034   used a repeated-measures two-way ANOVA across recording sites, with object/non-object contour

L035   and scrambled/not scrambled as factors.

L036

L037   **Data availability**

L038   Data will be available upon publication of the paper.

L039

L040   **Code availability**

L041   Custom code will be available upon publication of the paper.

L042

L043   **References**

L044   64.   Brainard, D. H. The Psychophysics Toolbox. *Spat Vis* **10**, 433–436 (1997).

65. Pelli, D. G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* **10**, 437–442 (1997).

66. Nestares, O. & Heeger, D. J. Robust multiresolution alignment of MRI brain volumes. *Magn Reson Med* **43**, 705–715 (2000).

67. Smith, S. M. *et al.* Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* **23 Suppl 1**, S208-19 (2004).

68. Yushkevich, P. A. *et al.* User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *NeuroImage* **31**, 1116–1128 (2006).

69. Wandell, B. A., Chial, S. & Backus, B. T. Visualization and measurement of the cortical surface. *J Cogn Neurosci* **12**, 739–752 (2000).

70. Harvey, B. M. & Dumoulin, S. O. 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).

71. Wandell, B. A., Dumoulin, S. O. & Brewer, A. A. Visual field maps in human cortex. *Neuron* **56**, 366–383 (2007).

72. DeYoe, E. A. *et al.* Mapping striate and extrastriate visual areas in human cerebral cortex. *Proc Natl Acad Sci U S A* **93**, 2382–2386 (1996).

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

74. Sereno, M. I. *et al.* Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* **268**, 889–893 (1995).

75. Friston, K. J. *et al.* Analysis of fMRI time-series revisited. *NeuroImage* **2**, 45–53 (1995).

L068   76.   Friston, K. J. *et al.* Event-related fMRI: characterizing differential responses. *NeuroImage* **7**,

L069   30–40 (1998).

L070   77.   Zuiderbaan, W., Harvey, B. M. & Dumoulin, S. O. Modeling center-surround configurations

L071   in population receptive fields using fMRI. *Journal of Vision* **12**, 10 (2012).

L072   78.   Dumoulin, S. O. & Knapen, T. How Visual Cortical Organization Is Altered by Ophthalmologic

L073   and Neurologic Disorders. *Annual Review of Vision Science* **4**, 357–379 (2018).

L074   79.   Bex, P. J. & Makous, W. Spatial frequency, phase, and the contrast of natural images. *J Opt

L075   Soc Am A Opt Image Sci Vis* **19**, 1096–1106 (2002).

L076   80.   Boynton, G. M., Demb, J. B., Glover, G. H. & Heeger, D. J. Neuronal basis of contrast

L077   discrimination. *Vision Res* **39**, 257–269 (1999).

L078   81.   Self, M. W., Kooijmans, R. N., Super, H., Lamme, V. A. & Roelfsema, P. R. Different glutamate

L079   receptors convey feedforward and recurrent processing in macaque V1. *Proc Natl Acad Sci

L080   U S A* **109**, 11031–11036 (2012).

L081   82.   Super, H. & Roelfsema, P. R. Chronic multiunit recordings in behaving animals: advantages

L082   and limitations. *Prog Brain Res* **147**, 263–282 (2005).

L083   83.   Cohen, M. R. & Maunsell, J. H. Attention improves performance primarily by reducing

L084   interneuronal correlations. *Nature Neuroscience* **12**, 1594–1600 (2009).

L085   84.   Palmer, C., Cheng, S. Y. & Seidemann, E. Linking neuronal and behavioral performance in a

L086   reaction-time visual detection task. *J Neurosci* **27**, 8122–8137 (2007).

L087   85.   Trautmann, E. M. *et al.* Accurate Estimation of Neural Population Dynamics without Spike

L088   Sorting. *Neuron* **103**, 292-308 e4 (2019).

L089   86.   Motter, B. C. Focal attention produces spatially selective processing in visual cortical areas

L090   V1, V2, and V4 in the presence of competing stimuli. *J Neurophysiol* **70**, 909–919 (1993).

L091   87.   Lescroart, M. D., Stansbury, D. E. & Gallant, J. L. Fourier power, subjective distance, and

L092   object categories all provide plausible models of BOLD responses in scene-selective visual

L093   areas. *Front Comput Neurosci* **9**, 135 (2015).

L094   88.   Knijnenburg, T. A., Wessels, L. F. A., Reinders, M. J. T. & Shmulevich, I. Fewer permutations,

L095   more accurate P-values. *Bioinformatics* **25**, i161–i168 (2009).

L096   89.   van der Walt, S., Colbert, S. C. & Varoquaux, G. The NumPy array: A structure for efficient

L097   numerical computation. *Computing in Science and Engineering* **13**, 22–30 (2011).

L098   90.   Abadi, M. *et al.* TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed

L099   Systems. (2016).

L100   91.   Thirion, B. *et al.* Inverse retinotopy: inferring the visual content of images from brain

L101   activation patterns. *NeuroImage* **33**, 1104–1116 (2006).

L102   92.   Stanley, G. B., Li, F. F. & Dan, Y. Reconstruction of natural scenes from ensemble responses

L103   in the lateral geniculate nucleus. *J Neurosci* **19**, 8036–8042 (1999).

L104   93.   Kay, K. N., Naselaris, T., Prenger, R. J. & Gallant, J. L. Identifying natural images from human

L105   brain activity. *Nature* **452**, 352–355 (2008).

L106   94.   Miyawaki, Y. *et al.* Visual image reconstruction from human brain activity using a

L107   combination of multiscale local image decoders. *Neuron* **60**, 915–929 (2008).

L108   95.   Nishimoto, S. *et al.* Reconstructing visual experiences from brain activity evoked by natural

L109   movies. *Current Biology* **21**, 1641–1646 (2011).

L110   96.   Freeman, J. & Simoncelli, E. P. Metamers of the ventral stream. *Nature Neuroscience* **14**,

L111   1195–1201 (2011).

L112

L113