Article

Current Biology

The Segmentation of Proto-Objects in the Monkey Primary Visual Cortex

Highlights

- V1 represents proto-objects and, later, globally perceived figure-ground structure
- Representations of proto-figures are enhanced and proto-grounds suppressed
- Proto-ground suppression develops after learning and can predict performance
- Proto-object modulation is present in the absence of attention

Authors

Matthew W. Self, Danique Jeurissen, Anne F. van Ham, Bram van Vugt, Jasper Poort, Pieter R. Roelfsema

Correspondence

m.self@nin.knaw.nl

In Brief

Self et al. show that V1 neural activity initially represents proto-objects. After monkeys learn to discriminate complex shapes, proto-ground ground regions become suppressed, so that late V1 activity strongly resembles our global perception of figures and correlates with behavior.

Self et al., 2019, Current Biology *29*, 1019–1029
March 18, 2019
https://doi.org/10.1016/j.cub.2019.02.016

CellPress

Current Biology
Article

The Segmentation of Proto-Objects in the Monkey Primary Visual Cortex

Matthew W. Self,[1,6,7,*] Danique Jeurissen,[1,2,6] Anne F. van Ham,[1] Bram van Vugt,[1] Jasper Poort,[1,3] and Pieter R. Roelfsema[1,4,5]

[1]Department of Vision & Cognition, Netherlands Institute for Neuroscience, Meibergdreef 47, 1105 BA Amsterdam, the Netherlands
[2]Department of Neuroscience, Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY 10032, USA
[3]Department of Psychology, University of Cambridge, Cambridge CB2 3EB, UK
[4]Department of Integrative Neurophysiology, Center for Neurogenomics and Cognitive Research, VU University, De Boelelaan 1085, 1081HV Amsterdam, the Netherlands
[5]Psychiatry department, Academic Medical Center, Postbus 22660, 1100DD Amsterdam, the Netherlands
[6]These authors contributed equally
[7]Lead Contact
*Correspondence: m.self@nin.knaw.nl
https://doi.org/10.1016/j.cub.2019.02.016

SUMMARY

During visual perception, the brain enhances the representations of image regions that belong to figures and suppresses those that belong to the background. Natural images contain many regions that initially appear to be part of a figure when analyzed locally (proto-objects) but are actually part of the background if the whole image is considered. These proto-grounds must be correctly assigned to the background to allow correct shape identification and guide behavior. To understand how the brain resolves this conflict between local and global processing, we recorded neuronal activity from the primary visual cortex (V1) of macaque monkeys while they discriminated between n/u shapes that have a central proto-ground region. We studied the fine-grained spatiotemporal profile of neural activity evoked by the n/u shape and found that neural representation of the object proceeded from a coarse-to-fine resolution. Approximately 100 ms after the stimulus onset, the representation of the proto-ground region was enhanced together with the rest of the n/u surface, but after ~115 ms, the proto-ground was suppressed back to the level of the background. Suppression of the proto-ground was only present in animals that had been trained to perform the shape-discrimination task, and it predicted the choice of the animal on a trial-by-trial basis. Attention enhanced figure-ground modulation, but it had no effect on the strength of proto-ground suppression. The results indicate that the accuracy of scene segmentation is sharpened by a suppressive process that resolves local ambiguities by assigning proto-grounds to the background.

INTRODUCTION

Proto-objects are the initial, rapidly formed neural representations of regions of the visual scene which, at a local level, resemble objects [1, 2]. Proto-object representations are an initial guess or sketch [3] by the visual system of the structure of the scene, which are presumably based on rapid local interactions between neurons in the early visual system. These representations have been proposed to be the neural substrate upon which attention acts [1, 4, 5], selecting particular proto-object representations for further processing. A region of the visual scene can also be considered to be a proto-object if it is extracted by the visual system as a region of interest. Such proto-object regions usually possess local Gestalt properties typical of natural objects (e.g., they are convex and are enclosed by borders) (Figure 1A). Proto-object regions can be divided into those that truly belong to foreground objects when the global scene is considered (proto-figures; e.g., the head of the snake in Figure 1A) and those that actually belong to the background (proto-grounds; e.g., the proto-object formed by the snake and the branch indicated in blue in Figure 1A). The neural mechanisms by which the brain selects and groups proto-figures into objects, while excluding proto-grounds, are currently unknown. The primary visual cortex (V1) may play a critical role in correctly assigning proto-objects to figures or backgrounds, as the activity of neurons in V1 is enhanced in regions perceived to be figures and suppressed on backgrounds [6–8], an effect known as figure-ground modulation (FGM). FGM is thought to arise through feedback from extrastriate visual areas [9], which reorganizes V1 activity to reflect the current global perceptual interpretation of the scene. However, previous neurophysiological studies of figure-ground segregation have used simple square stimuli that do not contain ambiguous proto-object regions [6, 8, 10, 11], whereas psychophysical studies revealed that the status of proto-ground regions can be ambiguous [12]. A critical outstanding question is therefore whether V1 represents the preliminary proto-object stage of form processing or a final perceptual stage in which proto-objects are correctly assigned to figure and ground. Alternatively, a previous computational

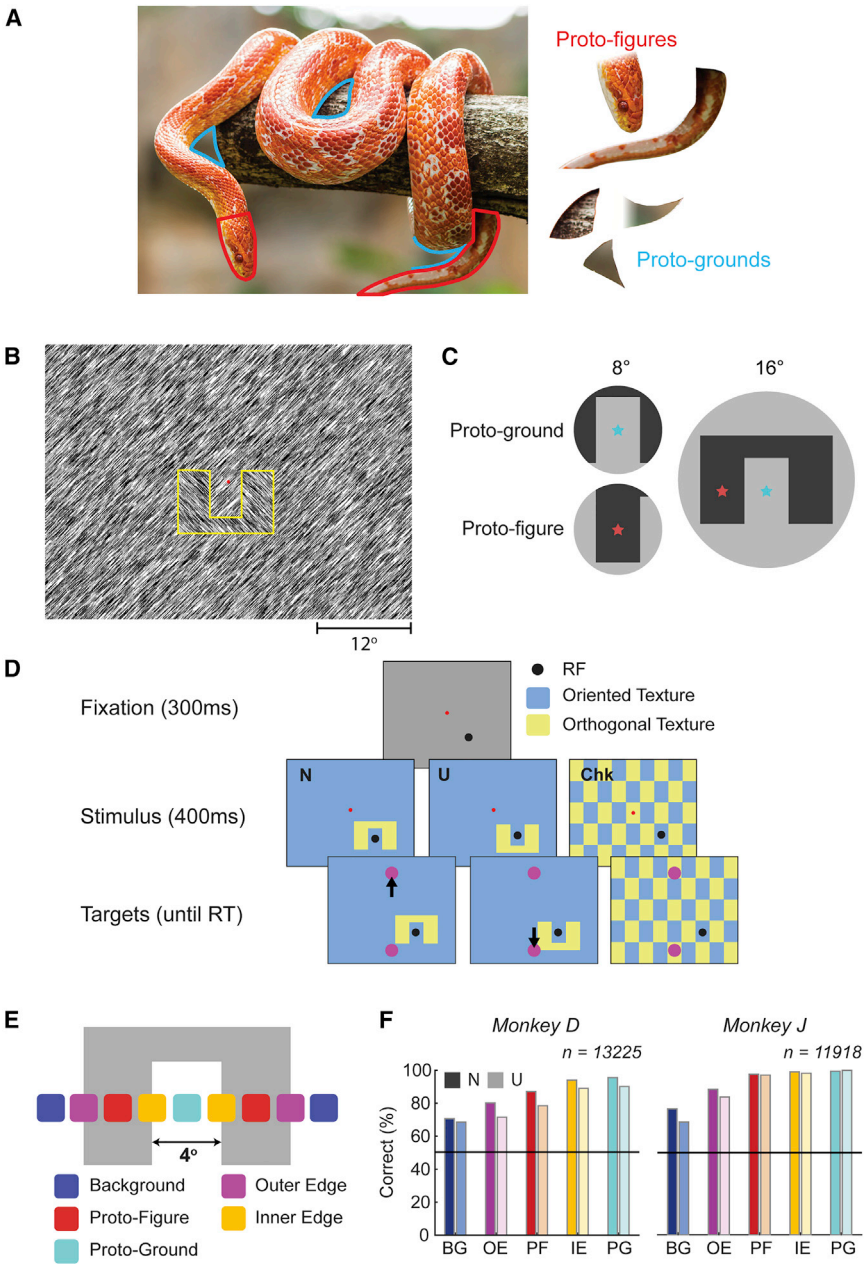

Figure 1. Proto-Object Stimuli and Task

(A) Natural scenes contain many boundaries that enclose image regions which, when analyzed locally, suggest the presence of an object. The regions illustrated in red show proto-figures, image regions that belong to objects. The regions enclosed in blue show example proto-grounds, which may first look like objects, but are actually background regions. The correct assignment of proto-figures and proto-grounds is critical for the recognition (and localization) of objects.

(B) A screenshot showing a full-screen texture with a texture-defined u form. A small red fixation point is visible in the center of the display. The u form is outlined for clarity here (the yellow lines were not shown to the monkey).

(C) When viewed through an aperture with a diameter of 8° (left), it is impossible to determine whether the blue or red asterisk is located on the figure. The true figure-ground assignment only becomes apparent when the stimulus is viewed through a larger aperture (right). Note that the monkeys always viewed full-screen textures and that they never saw the texture through an aperture.

(D) Animals were trained to discriminate between n's and u's. They started a trial by directing their gaze to the red fixation dot. After 300 ms, the full-screen texture appeared, containing either an n, u, or a checkerboard. After another 400 ms, the fixation dot was extinguished, and the animal was required to make an eye movement to one of the two magenta targets (the association between the shape and target was opposite for the two monkeys) or to maintain fixation in checkerboard trials.

(E) We compared activity elicited by the different parts of the texture-defined shapes by shifting the n/u form or checkerboard pattern relative to the V1 receptive fields across trials. Note that most conditions (except the proto-ground) occurred twice, on the left and right side of the figure. The checkerboard stimulus provides a measure for the activity elicited by boundaries and the homogeneous textures inside the checks, in the absence of clear figure-ground organization.

(F) Accuracy in the n/u discrimination task. The colors indicate which part of the stimulus the RF was centered on in accordance with the scheme in Figure 1E. Performance for all conditions was significantly greater than chance (all $p < 0.001$, binomial test). The accuracy was lowest for the background and outer-edge conditions because the shapes appeared on average at a larger eccentricity, making them harder to discriminate (see Figures S2B and S2C for the accuracy for all individual figure locations).

See also Figures S1 and S2.

model of figure-ground segregation [13] predicted that different levels of representation may occur sequentially at different time points of V1's response. The model predicts that, initially, activity elicited by all proto-objects would be enhanced, whereas the response elicited by proto-grounds would only be suppressed at a later point in time.

In this study, we examined the response of V1 to large, texture-defined n or u shapes (Figure 1B) that contain a proto-ground region that looks like a foreground object when analyzed at an intermediate spatial scale, but actually belongs to the background (Figure 1C). Such proto-ground regions cannot be distinguished from proto-figures through local mechanisms (Figure S1) and can be perceived as having their own shape. Previous studies demonstrated that the figure-ground status of proto-objects may, under some conditions, even remain ambiguous [12]. Using these complex forms, we investigated whether FGM in V1 reflects local proto-object structure or the global assignment of figure and ground. We find that both stages of processing are sequentially represented in V1. The spatial pattern of neural activity proceeds from an initial, coarse figure-ground segregation in which activity on both proto-figures and proto-grounds is enhanced, through to a later stage in which the representation

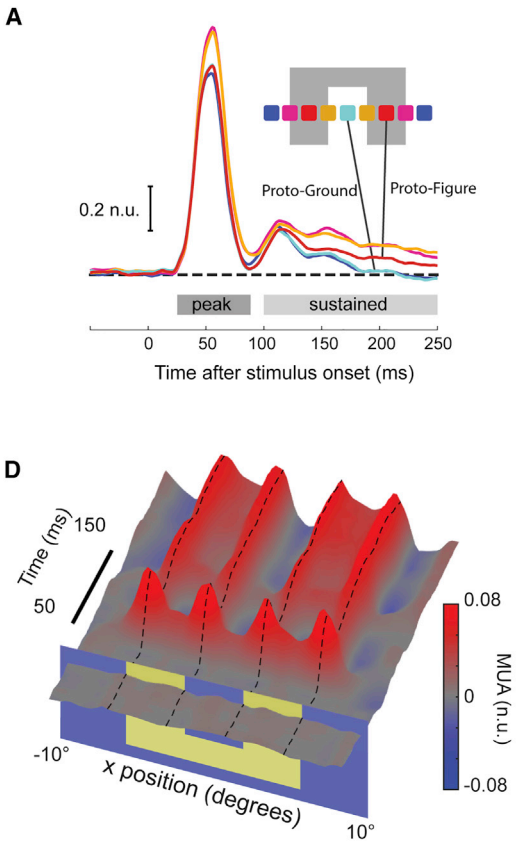

Figure 2. V1 Activity Elicited by the n/u Stimuli during the Shape-Discrimination Task

(A) The response evoked by the five different image regions at an example V1 recording site. The edges (magenta and orange lines) elicited extra V1 activity during the peak response (30–80 ms), whereas activity elicited by the proto-figure (red line) became stronger than that elicited by the proto-ground (cyan line) at a later point in time (100–250 ms).

(B) The x axis shows the responses during the peak phase from all V1 electrodes elicited by the conditions without boundaries in the RF (background, proto-figure, and proto-ground), and the y axis shows responses from inner and outer edges. Filled symbols, recording sites for which the difference was significant (t test, $p < 0.05$). Almost all sites responded more strongly to boundaries.

(C) V1 responses elicited by the proto-figure (x axis) and the proto-ground (y axis) condition in the sustained (100–250 ms) time window. The proto-figure elicited stronger V1 activity.

(D) We used the RF scatter (Figures S2 and S5) to create spatiotemporal maps of V1 figure-ground modulation. The 3D surface shows the enhancement (red colors) or suppression (blue colors) relative to the background. The stimulus onset is marked by the blue-yellow surface, which also marks the position of the n/u stimulus. The bar along the y axis shows the time scale in milliseconds.

(E) Spatial profile of the average activity in the 100- to 250-ms time window. The gray areas mark the location of the proto-figures.

See also Figures S3, S4 and S5.

of the proto-ground is suppressed back to the level of the background, so that the resulting spatial pattern of neural activity strongly resembles our enhanced perception of figure surfaces [14–16]. To probe the mechanisms of proto-object assignment, we studied whether proto-figures and proto-grounds can be segregated when the shape is unfamiliar to the animal and when the animal directs its attention elsewhere. We find that figure-ground modulation is present in naive animals but that the suppression of the proto-ground was only present when the animals were trained to recognize the shapes. Attention enhanced figure-ground modulation, but it had no effect on the suppression of the proto-ground region.

RESULTS

We trained two macaque monkeys to perform a shape-discrimination task. The monkeys first directed their gaze to a red fixation point on a gray background, and after 300 ms, we presented a full-screen texture, which contained either a texture-defined n or u form or a control checkerboard with ambiguous figure-ground organization (Figures 1B and 1D). We chose n and u shapes for this task because they contain proto-object regions which appear to be figural at the scale of the receptive fields of neurons in the early and mid-level visual cortex (Figure 1C and Figure S1). We refer to the legs of the n/u as proto-figures and the central background region as proto-ground because these regions will ultimately be assigned to figure and ground in perception, respectively. The animals reported the form that they perceived by making a saccadic eye movement to one of two circular targets (Figure 1D). We examined multi-unit activity (MUA) in area V1 using chronically implanted electrode arrays (Utah probes). Across trials, we positioned the n/u form at different locations relative to the neurons' receptive fields (RFs) to sample activity elicited by the figure, the ground, the edges between figure and ground and the proto-ground (there were 5 main conditions, Figures 1E and S2A). The accuracy of both monkeys was well above chance at all positions (average = 89.3% for monkey D and 88.6% for monkey J, Figure 1F and Figure S2B), although monkeys did significantly better when the shape was presented close to fixation (Figure S2C; χ^2 test on effect of condition: monkey D, $\chi_4^2 = 385$; monkey J, $\chi_4^2 = 1116$; both $p < 0.001$). We will first describe the V1 activity profile when the monkeys had become proficient in this task, and we will then compare it to the activity before learning to discriminate between the n and u.

Figure 2A shows multi-unit activity from an example V1 recording site for each of the 5 conditions (three further examples are shown in Figures S3A–S3C). In this analysis, we averaged across stimuli with both orientations of the texture elements, thereby ensuring that the same texture elements were present at all locations so that the differences in activity could not be caused by a difference in orientation or position of the

texture elements in the neurons' receptive fields (analyses in which the different orientations were held separate are shown in Figure S3D). Neural activity was modulated by the position of the figure in two successive phases. In the early peak phase (30–80 ms), responses were significantly higher when a textured boundary was present in the RF than they were in the non-boundary conditions (yellow and pink curves in Figure 2A) (two-sample t test, $t_{2860(trials)} = 18.4$, $p < 0.001$). Significant early boundary modulation was present at 80% of recording sites (two-sample t test, $p < 0.05$, for 75 out of 94 sites, Figure 2B) as well as when we averaged across the population of recording sites (paired t test: monkey D, $t_{55} = 11.9$; monkey J, $t_{37} = 9.6$; both monkeys, $p < 0.001$). The early boundary modulation has been described previously [9, 17–19] and is thought to arise through inhibitory interactions between neurons that are tuned to the same orientation in the superficial layers of V1 [9]. This inhibition is strongest in image regions with a homogeneous orientation, and it is partly released at boundaries where the different orientations abut.

Of particular interest to this study is that responses at the example site were significantly higher on the proto-figure than on the proto-ground (two-sample t test, $t_{985(trials)} = 4.9$, $p < 0.001$) in a later phase of the response when activity reached a stable level (100–250 ms) (red versus cyan curves in Figure 2A). We will refer here to this late modulation as proto-object modulation (POM) to distinguish it from standard FGM. In our analysis of the population of V1 recording sites, we normalized activity from individual recording sites by subtracting the pre-stimulus activity and dividing by the peak response. We will express effect sizes in normalized units (n.u., Figure 2A), which represent fractions of the peak response. The average level of POM across the population was 0.034 n.u., or 3.4% of the peak response. This may seem to be a small effect, but in the time period of 100–250 ms, sustained neural responses had fallen to a level of ~0.1 n.u. Expressed as a percentage change, neural responses on the proto-figure were 38.3% higher than on the proto-ground, a substantial modulation of the activity. The relative increase of activity on the proto-figure was also highly consistent across the population of recording sites in both monkeys. POM was significant in 71 out of 94 sites (two-sample t test, $p < 0.05$) and was also significant across the population in each monkey individually (paired t test: monkey D, $t_{55} = 30.2$; monkey J, $t_{37} = 8.8$; both monkeys, $p < 0.001$, Figure 2C). The enhancement of the V1 representation of proto-figures relative to that of proto-grounds was not related to small differences in the position of the eye around the fixation point (Figure S4).

We examined the spatial properties of POM in more detail by taking advantage of the scatter of RF positions (Figures S5A–S5D) to produce interpolated maps of the modulation of the neuronal activity (Figure 2D). These maps show the change in V1 activity relative to the background condition (see Figure S5E for details). During the early phase, clear peaks were visible at the boundaries of the n/u figure. At later time points, the figure surface gradually became filled in with increased activity while responses on the proto-ground region were reduced back down to the level of the background leading to a neural segregation of proto-figures from proto-ground. The segregation of figure and background during the sustained-response phase was remarkably sharp (Figure 2E) considering that this profile is unavoidably blurred by slight differences in gaze across trials (Figure S4) and imprecisions in our assessment of the RF positions.

Enhancement or Suppression?

The difference in activity elicited by the proto-figure and proto-ground could be due to enhanced activity on the proto-figure, suppression of the proto-ground, or both. To distinguish between these alternatives, we created a matched checkerboard control for each n/u stimulus (Figure 1D). Checkerboards have clear boundaries that result in perceptual segregation, but there is no unambiguous assignment of particular checks to figure or background, making them well suited as a reference stimulus. A matched checkerboard was constructed for each stimulus position so that the boundaries of the checkerboard coincided with the boundaries of the n/u forms (Figure 3A). To isolate the figure-ground assignment signal, we subtracted the checkerboard response from that elicited by the n/u shapes. We then used a passive fixation task (the monkeys maintained gaze on the fixation point for 400 ms for all stimuli) to avoid task-induced differences in the processing of checkerboards and n/u forms. Figure 3B illustrates the spatial profile of V1 activity elicited by n/u forms and checkerboards. Compared to the representation of the checkerboard, the representation of proto-figures was enhanced (red areas in Figure 3C), whereas that of the proto-ground was suppressed (blue areas in Figure 3C). The strength of these two effects was not correlated across recording sites (monkey D: Pearson's $r = -0.03$, monkey J: $r = 0.04$; both $p > 0.05$), suggesting that enhancement and suppression may arise through separate processes [7]. The time course of Figure 3D revealed an early phase of enhanced V1 activity for both proto-figures and proto-ground relative to their checkerboard controls at around 100 ms (proto-figure latency = 98.8 ms, proto-ground latency = 91.0 ms, black arrow in Figure 3D). This phase coincides with the latency of FGM in simple figure-ground tasks [9, 18] and represents a coarse form of figure-ground segregation that apparently does not yet distinguish between proto-figures and proto-grounds. The difference in response elicited by proto-figures and proto-grounds first arose after 115 ms (paired t test latency analysis, see STAR Methods; Figure 3D), significantly later than the latency of the initial, coarse enhancement phase (bootstrap test, see STAR Methods, $p = 0.01$ for both proto-figure and proto-ground, compare the vertical black dashed line to the blue and red lines in Figure 3D). These results imply that the segregation of n and u figures proceeds in two stages. During an early phase, V1 activity highlights both proto-grounds and proto-figures. Slightly later, the activity elicited by the proto-ground is suppressed relative to the activity elicited by the proto-figure.

Relationship between V1 Activity and Behavior

We computed choice probabilities to investigate how the modulation of V1 neural activity correlates with the performance of the animals on the shape-discrimination task. Choice probability is larger than 0.5 if increased V1 MUA predicts that the monkey will get it right, smaller than 0.5 if it predicts an error, and equal to 0.5 if V1 activity is unrelated to the choice. Whereas previous studies usually focused on the choice probability for one feature [20, 21], the small V1 receptive fields allowed us to determine which parts of the V1 representation best predict performance. Specifically, we computed the average choice probability

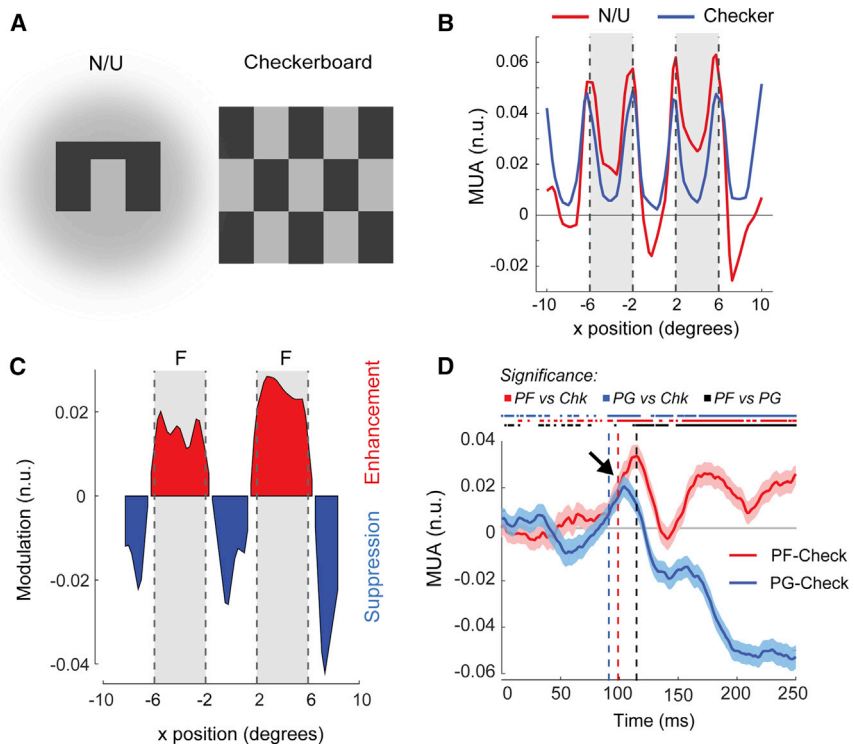

Figure 3. Comparison between V1 Activity Elicited by the n/u Stimulus and the Checkerboard

(A) We used a texture-defined checkerboard as a reference stimulus because it has similar boundaries to the n/u shapes, but it lacks clear figure-ground organization.

(B) The spatial profile of activity (100–250 ms) in V1 for n/u stimuli and checkerboards. Zero on the y axis indicates the average activity on the background of the n/u stimulus.

(C) The difference between activity evoked by n/u stimuli and checkerboards revealed regions of figure enhancement and ground suppression.

(D) The time course of figure enhancement and ground suppression in V1. The arrow marks an early phase during which responses elicited by proto-figures and proto-grounds were both enhanced relative to those elicited by the checkerboard. The blue, red, and black checks above the graph mark samples for which there was a significant difference between activity elicited by the proto-figure and the checkerboard (red, paired t test, $p < 0.05$), the proto-ground and the checkerboard (blue), and the proto-figure and proto-ground (black). The latencies of these effects were defined as the first of 11 consecutive significant samples (dashed lines). The shaded region indicates ± 1 SEM.

separately for trials with RFs on the background, one of the edges, the proto-figure, or proto-ground (Figure 4A). Choice probability depended on the RF location (Kruskal-Wallis test, $\chi_4^2 = 11.3$, $p = 0.02$). Suppression of the proto-ground best predicted the monkey's performance: the accuracy was higher if the suppression of the proto-ground was stronger (Figure 4B, median choice probability = 0.47, $p = 0.02$, Wilcoxon signed-rank test). Accordingly, V1 activity evoked by the proto-ground was higher on error trials (Figure 4C, $p = 0.002$, t test). We also observed a weak influence of the outer edge (median choice probability = 0.51, $p = 0.004$, Wilcoxon signed-rank test). V1 activity elicited by the other regions did not predict performance (Figure 4B, left; all $p > 0.05$). Does this mean that the relative enhancement of activity on the proto-figure was irrelevant for the monkey's performance? The only difference between an n and u stimulus is the position of the horizontal segment that links the two legs of the figure (the discriminant segment, Figure 4A). In monkey J, we therefore also examined the choice probability for the discriminant segment. We moved the n/u figures so that either the discriminant segment (DF in Figure 4A) or the corresponding ground region (DG) was in the RF. The monkey's accuracy was similar to that in the other conditions (Figure S6A). Activity elicited by DF also predicted the monkey's accuracy with median choice probabilities of 0.53 (Wilcoxon signed-rank test, $p = 0.001$), but the effect of DG activity was not significant (choice probability = 0.48, $p = 0.13$) (Figure 4B, right). Indeed, on correct trials but not on erroneous trials, V1 activity elicited by DF was stronger than that elicited by DG (Figure 4D) (correct trials, t test, $p < 0.01$; error trials, $p = 0.9$), an effect that was also reliable when we equated the numbers of correct and error trials (Figure S6B). Thus, figure-ground modulation was absent on error trials, but it was not reversed, which suggests that errors were associated with a failure to correctly segregate the discriminant segment rather than perceiving an illusory segment at the wrong position. In summary, the analysis of choice probability indicated that accurate performance on this shape-discrimination task is correlated with the suppression of the proto-ground and the enhancement of the discriminant segment.

The Effect of Training on Proto-Object Modulation

Did POM arise through extensive training with the n/u forms? Prior to training the animals on the shape-discrimination task, we recorded neural activity in V1 during passive fixation. monkey D had not been trained on any figure-ground task before, whereas monkey J had performed figure-ground tasks with simple square shapes among other tasks (STAR Methods), but monkey J had never seen the n/u shapes before or performed shape-discrimination tasks. The training procedure took 36 and 65 days in monkeys D and J, respectively. After training, the monkeys carried out the passive fixation task once more, so that we could compare V1 activity before and after training during an identical task.

Even before training, V1 neurons discriminated between figure and background, because the proto-figure elicited stronger activity than did the proto-ground (Figure 5A, t tests, both animals, $p < 0.001$). Comparisons with the checkerboard control (Figure 5B, left) revealed that in this phase, POM was driven by enhancement of the proto-figure (t test, $t_{93} = 15.1$, $p < 0.001$) without significant proto-ground suppression (t test, $t_{93} = -2.2$, $p = 0.14$). Training to distinguish between the texture-defined n/u forms caused changes in the V1 representation (Figure 5B, right). The strongest training effect was an increased suppression of the proto-ground (arrow in Figure 5B), down to the level

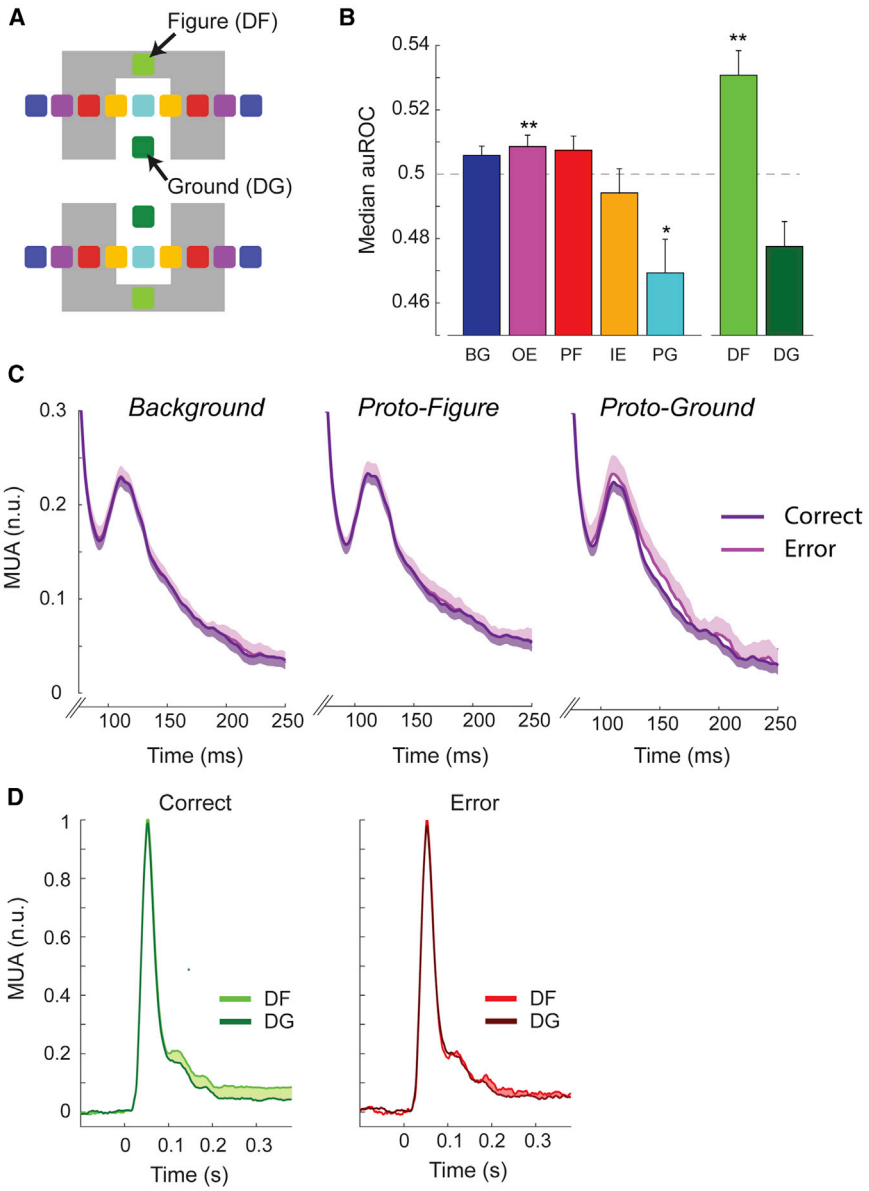

Figure 4. Correlation between V1 Activity and the Animal's Choice

(A) To discriminate between the n and u, the monkeys needed to locate the discriminant segment, connecting the two legs of the figure. We studied V1 responses elicited by this segment in monkey J by placing it (or the equivalent location of the background) in the RF. This created two extra conditions: discriminant figure (DF) and discriminant ground (DG).

(B) Choice probabilities. The choice probability is higher (lower) than 0.5 if increased (decreased) V1 activity predicts a correct choice. The data from the 5 main conditions (BG, OE, PF, IE, and PG) come from both animals; the discriminant segment (DF and DG) was only tested in monkey J. Asterisks mark conditions in which the median choice probability was significantly different from 0.5 (Wilcoxon signed-rank test, *$p < 0.05$, **$p < 0.01$). Error bars indicate 1 SEM.

(C) V1 activity elicited by three different parts of the n/u stimulus on correct and error trials. Responses evoked by the proto-ground were higher on error trials, implying a lack of suppression. Shaded region indicates +1 SEM (error trials) and −1 SEM (correct trials).

(D) Activity elicited by the discriminant segment of the figure and ground on correct and error trials. See also Figure S6.

of the V1 background response outside the figure (BG in Figure 5C). This effect was also clear without subtracting the checkerboard control (Figures S7A and S7B) and occurred in both monkeys (Figure S7C). For statistical analysis, we subtracted the checkerboard response and applied a repeated-measures two-way ANOVA with training and receptive field position as factors (2 and 5 levels, respectively, Figure 5C). The interaction between receptive field position and training was significant ($F_{2.9,\ 266.6} = 20.5$, $p < 0.001$). Post hoc t tests revealed that training suppressed responses elicited by the proto-ground ($t_{93} = -7.6$, $p < 0.001$) and background ($t_{93} = -4$, $p = 0.001$); furthermore, the effect of training was significantly stronger on the proto-ground region than on the background ($t_{93} = -5.1$, $p < 0.001$). Thus, the main effect of training was a more efficient suppression of background regions with a particularly strong effect on the proto-ground.

The Effect of Attention on Proto-Object Modulation

In the previous section, we used a passive-fixation task to compare pre- and post-learning neural data. The low task demand of passive fixation leaves open the possibility that the monkeys still engaged in the shape-discrimination task while fixating and that the observed increase in proto-ground suppression might have been the result of attentional selection of the shape. After collecting the data in the passive-fixation task, we determined whether the suppression of the proto-ground was the result of attentional selection by retraining the animals to ignore the textured shape and to perform the shape-discrimination task on a color-defined n/u figure in the opposite hemifield (Figures 6A and 6B). The n/u shape of the texture-defined figure that had to be ignored was uncorrelated with the shape of the color-defined shape so that paying attention to former shape would have hampered performance. To ensure that the monkeys did not switch back to judging the texture-defined figure, we did not interleave the different tasks on different days, but instead compared the data in which the animals attended to the color-defined figure to the previously recorded data in which the animals attended the texture-defined figure. The monkeys were able to successfully ignore the textured n/u shape, as evidenced by the above-chance performance on the colored-shape task (the average accuracy of monkey D was 74.4% and that of monkey J was 88.6%) and the lack of any relationship between the

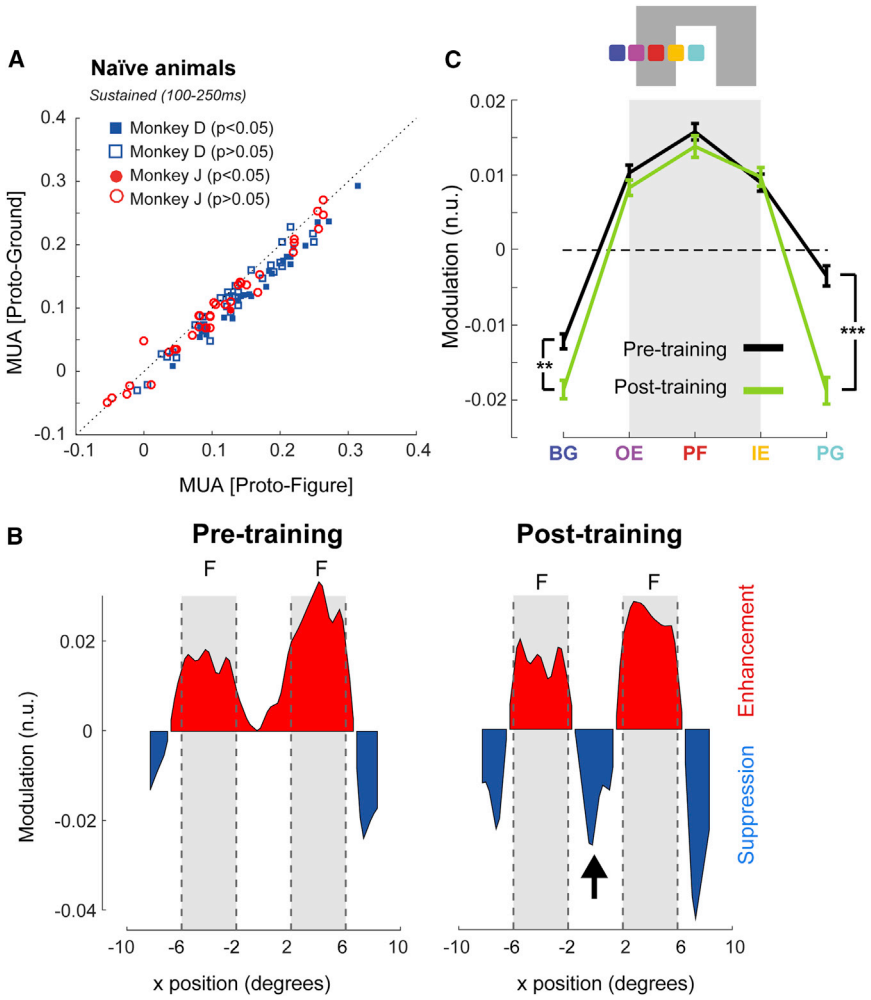

Figure 5. The Effects of Shape-Discrimination Training on the V1 Representation of the Texture-Defined Figure

(A) Responses elicited by the proto-figure and proto-ground in naive animals, showing significant POM before training.

(B) Spatial profiles of the response difference between the n/u figures and checkerboards in naive and trained animals (100–250 ms). The arrow marks the increase in proto-ground suppression due to training.

(C) Response differences between the n/u and checkerboard grouped into the five main conditions for statistical analysis. Training significantly reduced responses evoked by the proto-ground and the background. Error bars indicate ± 1 SEM; asterisks mark significant differences between conditions (post hoc t tests, Bonferroni corrected, $^{**}p < 0.01$, $^{***}p < 0.001$).

See also Figure S7.

position of the textured n/u shape and correct performance on the colored n/u (compare Figure 6C with Figure 1F). Despite the withdrawal of attention from the textured shapes, we still observed significant proto-object modulation in both animals (Figure 6D, t test, $p < 0.001$, both monkeys). Note that only trials in which the animal correctly determined the shape of the colored n/u were included in the analysis. To examine the effect of attention on the different compartments of the shape, we compared neural responses when the monkeys were actively attending to the textured figure to those when they ignored it (Figures 6E and 6F). Withdrawal of attention influenced the representation of the texture-defined form, as revealed by a significant interaction between receptive field position and attention in a two-way ANOVA ($F_{1.5,\ 137.3} = 23.1$, $p < 0.001$, Figure 6E). When attention was directed toward the figure, the activity for the proto-figure and boundary conditions increased. However, attention caused a strong and significant reduction in activity on the background (t test, $p < 0.001$, Bonferroni corrected). Remarkably, this reduced background activity did not extend into the proto-ground region; indeed, we found no significant attention-related changes in activity on the proto-ground (t test, $p = 0.58$, Bonferroni corrected). The results suggest that attention can select the entire figure region and coarsely suppress nearby background regions while not having any effect on the more perceptually ambiguous proto-ground regions. This result implies that the increase in proto-ground suppression observed after learning must be due to the learning process itself, rather than covert attentional selection of the figure after learning.

DISCUSSION

The visual system implements the Gestalt rules of perceptual organization to group together all the image regions that belong to a spatially extended object and to segregate them from the background [15, 22–24]. Local image regions may initially look like object parts, but they may actually belong to the background when the global context is taken into account. A previous study by Kim and Feldman [12], using shapes containing a convex, highly enclosed proto-ground region, suggested that such proto-grounds may not always be unambiguously assigned to the background in the final percept, even though correct figure-ground assignment is important for shape perception and for the guidance of behavior. In their study, a clever manipulation was used to establish to which of two abutting regions a contour was assigned. In the present study, we did not train monkeys to report the assignment of contours but to report the identity of a large n or u shape while measuring the responses of V1 neurons elicited by the interiors of proto-grounds and proto-figures. These shapes contain a proto-ground region which possesses many of the Gestalt cues that are normally diagnostic for objects in natural scenes (i.e., convex borders, enclosure, feature contrast), but which ultimately belongs to the background. We found that neural activity in V1 progressed through three phases within a trial: an early increase of activity at the locations of the texture-defined boundaries (Figure 7A), followed by a coarse form of figure-ground segregation at approximately 100 ms (Figure 7B), during

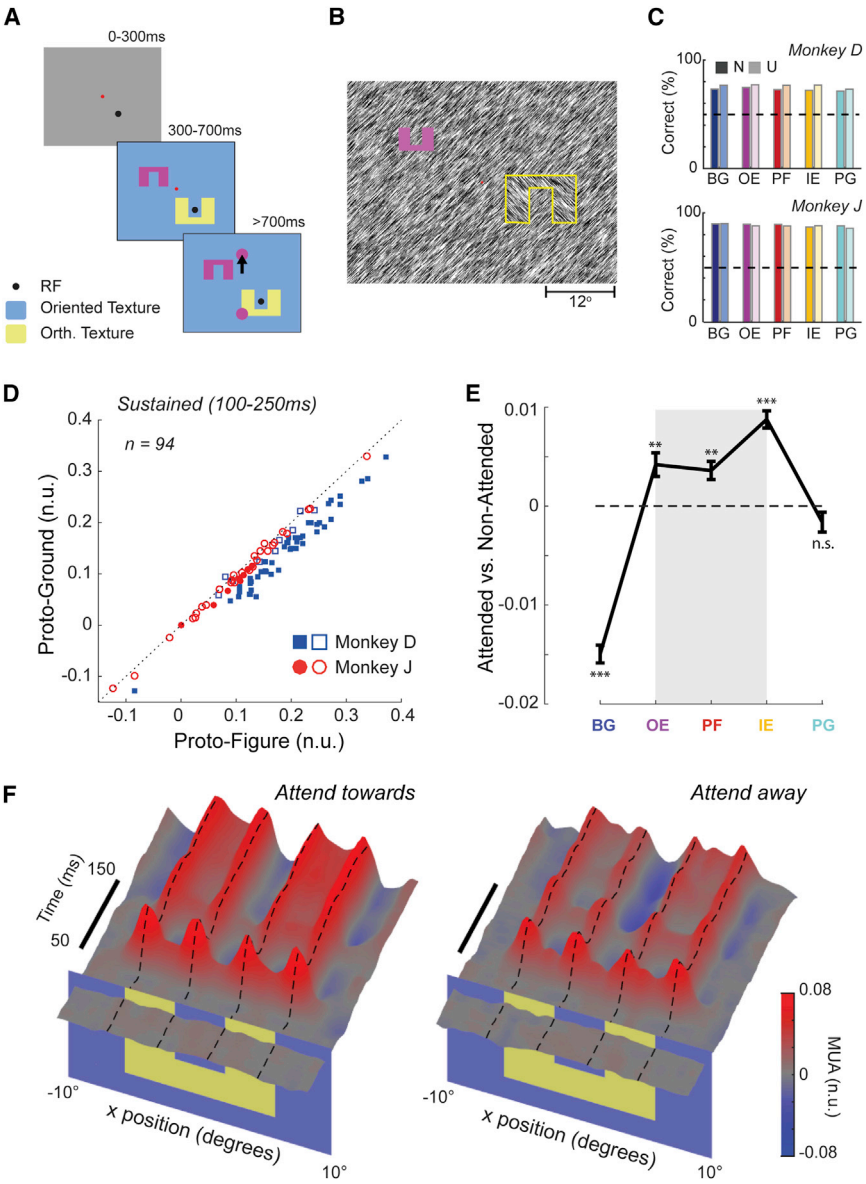

Figure 6. The Effect of Attention on Proto-Object Modulation

(A) Animals were trained to ignore the textured figure and perform the shape discrimination task on a magenta figure that was presented in the opposite visual quadrant.

(B) Screenshot of the stimulus showing the positioning of the magenta n/u figure. The texture-defined n/u figure is outlined here for clarity.

(C) The performance of the animal on the magenta shape-discrimination task. This performance was unrelated to the position of the texture-defined n/u (x axis).

(D) Neural activity during the sustained period (100–250 ms) for each electrode while attention was directed to the magenta n/u. Individually significant electrodes (t test, $p < 0.05$) are shown in filled symbols.

(E) The difference in activity for each of the shape compartments between days during which the animal was attending to the textured n/u and days during which the animal was ignoring it and attending the magenta n/u. Asterisks mark significant differences (Bonferroni-corrected paired t test, $^{**}p < 0.01$, $^{***}p < 0.001$). Error bars indicate +1 SEM.

(F) The spatiotemporal profile of figure-ground modulation when the animals were attending to the texture-defined n/u (left panel) or the magenta n/u (right panel) in the opposite visual quadrant. Conventions are as in Figure 2D.

which V1 activity elicited by both the proto-figure and the proto-ground was increased. Finally, after ~115 ms, V1 activity evoked by the proto-figure became enhanced, but that evoked by the proto-ground was suppressed, so that only texture elements of the n/u form were labeled with extra V1 activity, and the proto-ground was unambiguously assigned to the background (Figure 7C). These results indicate that V1 represents both an early proto-object phase and the later phase in which figure-ground assignment agrees with the global interpretation of the scene. Training on the shape-discrimination task increased the strength of proto-ground suppression (Figure 7D), and the importance of proto-ground suppression was supported by the finding that the degree of suppression was linked to variations in the monkeys' accuracy once they had learned the meaning of the two shapes. To isolate proto-figure enhancement and proto-ground suppression, we compared neural responses to the n/u shape to a matched checkerboard control. This procedure works under the assumption that the checkerboard is perceived without any clear figure-ground assignment, as it generally appears to human observers. However, it is possible that the monkeys perceived figure-ground structure within the checkerboard, and if the assignment of particular checks to figure or ground varied across trials, this would lead to an intermediate level of response on checkerboard trials and give a false impression of separate suppression and enhancement effects. This possibility appears unlikely, as the strength of proto-figure enhancement and proto-ground suppression was not correlated across channels, as would be expected if they were a single process, and the timing of the two effects also differed. Furthermore, if checks are sometimes perceived as figures and sometimes as ground, this could produce bimodal distributions of neural activity during the late modulation period. We examined the responses to checkerboards in this late period, but we did not observe any signs of bimodality (Hartigan's dip test, all channels, $p > 0.05$). The results therefore stand in support of two independent processes: an early one that enhances the response to all proto-objects, and a later one that suppresses proto-grounds.

Notably, our results settle two long-standing debates concerning FGM: first, that FGM may be due to spatial attention, and second, that it may only be present in over-trained animals [25, 26]. We found that FGM was present in naive animals who

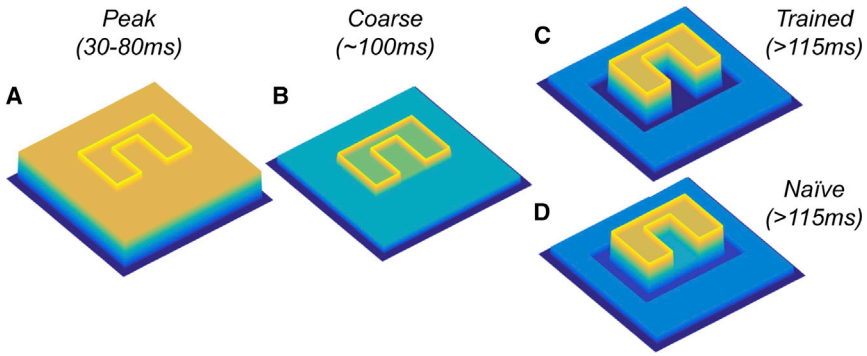

Figure 7. Graphical Summary of the Results

(A) During the peak phase, V1 neurons are strongly driven by onset of the texture elements in their RFs. The responses elicited by the texture boundaries are enhanced.

(B) During an intermediate phase (around 100 ms), V1 activity elicited by the proto-figure and proto-ground is also enhanced. The extra V1 activity provides information about the coarse location of the n/u.

(C) In trained animals, V1 activity elicited by the figure surface is enhanced, and activity elicited by the background and proto-ground is suppressed. The result is a robust neural segmentation of the figure surface in V1.

(D) In naive animals, the suppression of the proto-ground is incomplete, producing a weaker neural segmentation signal.

had not performed the shape-discrimination task, and it was present even in one animal (monkey D) who had never seen figure-ground textures. Furthermore, the modulation was observed in passively fixating animals and when attention was directed away from the figure, suggesting that it was independent of the task. Finally, the figure enhancement had an extremely precise spatial pattern, which could not be explained by a global focus of attention toward the figure location.

The Mechanisms That Assign Proto-Objects to Figure and Ground

The coarse-to-fine development of the spatial pattern of modulation in V1 closely resembles the results of hierarchical computational models of figure-ground modulation [13, 27]. In these models, the global figure-ground structure is first determined through competitive interactions between cells tuned for the same feature (e.g., orientation) at multiple levels of the visual hierarchy. At low levels of the hierarchy, this analysis of the scene cannot discriminate between proto-figures and proto-grounds; hence, proto-grounds are initially extracted as figures. However, at the higher levels, the receptive fields are large enough to discriminate figure from ground, and these areas use feature-specific feedback connections to overwrite the erroneously enhanced representation of the proto-ground in the lower layers at later time points, just as was observed in the present study (Figure 3D). These models provide a mechanism by which higher visual areas indicate which features in the scene are figural and which belong to the background.

However, models representing only the surface properties of objects, like texture, may run into difficulties if images contain overlapping objects. In these cases, information about which boundaries belong to which objects is critical to disambiguate the scene [28]. In these situations, neurons that code for border ownership may come into play. These neurons give a stronger response when a figure is located at a particular side of an edge in their receptive field, and they thereby encode which of the two image regions that abut at an edge "owns" it. Border-ownership-tuned neurons are present in large numbers in higher visual areas, such as the V2 and V4 [2, 5, 29, 30]. They may play a critical role in the assignment of proto-objects to figure and ground, because they have been shown to accurately represent edge assignment for n and u forms [30]. We therefore hypothesize that border-ownership neurons could inform V1 about the figure-ground assignment by sending excitatory feedback to the preferred figure side and suppressive feedback to the non-preferred side. Such a feedback signal would thereby help V1 to disambiguate between proto-figures and proto-grounds, enabling the correct labeling of the surface of a complex shape as a figure (M. Self et al., 2015, Soc. Neurosci., abstract).

Effects of Learning on Proto-Object Modulation

Previous studies in naive animals did not observe object-segregation signals in the early visual cortex [26, 31, 32]. We were therefore surprised to find weak, but significant, figure-ground modulation before the animals were trained in the task, even in monkey D, who had never been exposed to texture-defined stimuli. One possible explanation for this difference between findings is that we presented the n/u shape at multiple locations to assess the spatial profile of V1 responses. Accordingly, the monkeys gained many views of the figure, and they therefore may have started to recognize the presence of a shape [33].

Interestingly, the enhancement of the proto-figures could be dissociated from the suppression of the proto-ground. In naive animals, proto-figure enhancement was already present, but proto-ground suppression increased when the animals learned to recognize the shapes during a training process of several weeks (Figure 5B). These results differ from those of a recent study by Yan et al. [32] in which monkeys learned to detect contours composed of collinear line elements embedded in noise. In the study by Yan et al., learning both enhanced V1 responses elicited by the figural contour elements and suppressed the responses elicited by the background elements. By contrast, in the present study, the largest effect was an increase in the suppression of the representation of the proto-ground. It seems likely that this difference between results is related to the shape of the figures and the nature of the task. In our study, the monkeys learned to discriminate between two similar shapes, and learning may have selectively amplified the representation of the most informative parts of the shapes [34], such as the discriminant segment and the proto-ground region, which were the only regions that differed between the n and the u. In contrast, Yan et al. [32] trained monkeys to detect a string of collinear line elements, so that any enhancement of the figure or suppression of the background would improve the accuracy on the task.

The fact that proto-ground suppression increased when the animals learned to recognize the shapes is compatible with the

hypothesis that proto-ground suppression arises through feedback originating from shape-selective cells in higher visual areas. Many neurons in higher visual cortical areas are tuned to the shape of objects, and they retain their tuning when shapes are defined by a difference in the texture between foreground and background [35]. Their feedback to lower visual areas can help with the assignment of edges and image regions to figures, because shape familiarity is known to bias figure-ground perception [36]. We speculate that proficiency in the shape-discrimination task enhanced the tuning of neurons in higher visual areas for the n/u shapes, thereby increasing the suppressive feedback to the V1 representation of the ambiguous proto-ground region.

The level of suppression of the proto-ground predicted the monkeys' performance on the shape-discrimination task on a trial-by-trial basis (Figures 4A and 4B), raising the possibility that the spatial pattern of modulation in V1 contributes toward the animal's performance. This finding is in accordance with fMRI studies in humans, demonstrating that shape-discrimination training enhances activity in early visual areas [37]. However, we cannot infer a causal contribution of V1 to the discrimination of texture-defined shapes from correlative measures such as choice probability [38]. Nevertheless, we can exclude the idea that the correlation between V1 activity and behavior is due to non-specific effects, such as arousal, reward expectation, or stimulus eccentricity, because the choice probability was above 0.5 for the discriminant figural segment and below 0.5 for the proto-ground (Figure 4A, the monkeys' accuracy was similar for these regions; Figures 1E and S6A). Interestingly, the representation of the discriminant figure segment (DF in Figure 4) was only enhanced on correct trials, which suggests that it depended on the selection of the appropriate shape by higher areas [33]. On erroneous trials, it appears that these recurrent interactions failed, so the discriminant segment was not enhanced and the proto-ground not suppressed, which may have contributed to the sub-optimal performance on these trials.

Conclusions

The present results demonstrate that the spatial pattern of activity in V1 initially reflects the extraction of proto-objects from their background and, later, the correct grouping of proto-objects into spatially extended shapes. The end result is a precise spatial enhancement of the figural surface, which corresponds well to our enhanced perception of figures. The representation of all the texture elements within the complex shapes was enhanced in V1. This finding is difficult to explain by local processing within V1, but instead suggests that there are intricate recurrent interactions between V1, mid-level areas where neurons code for shape fragments and border ownership, and higher areas where neurons code for the overall shape of figural regions. Future studies could use related perceptual tasks and record from neurons at multiple levels of the visual cortical hierarchy to increase our understanding of how lower and higher cortical areas work together to extract meaningful and coherent object representations from crowded and complex visual scenes.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Training history and surgical details
  - Electrophysiology
  - Stimuli
  - Receptive Field Mapping
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Data Analysis
  - Analysis of eye position
  - Analysis of feature contrast
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found with this article online at https://doi.org/10.1016/j.cub.2019.02.016.

ACKNOWLEDGMENTS

P.R.R. was supported by NWO (ALW grant 823-02-010) and the European Union (ERC Grant Agreements 339490 "Cortic_al_gorithms" and 7202070 and 785907 "SGA1 and SGA2 of the Human Brain Project"). D.J. was supported by the Simons Foundation (414196). We thank Kor Brandsma and Anneke Ditewig for help with training the animals and surgical procedures.

AUTHOR CONTRIBUTIONS

M.W.S. and P.R.R. conceived of the research. M.W.S., P.R.R., and J.P. designed the paradigm. M.W.S., D.J., A.F.H., B.V., and J.P. recorded the data. M.W.S. analyzed the data with D.J. and J.P. M.W.S. and P.R.R. wrote the paper with D.J. and J.P.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 26, 2018
Revised: January 7, 2019
Accepted: February 5, 2019
Published: March 7, 2019

REFERENCES

1. Rensink, R.A. (2000). The dynamic representation of scenes. Vis. Cogn. *7*, 17–42.
2. von der Heydt, R. (2015). Figure-ground organization and the emergence of proto-objects in the visual cortex. Front. Psychol. *6*, 1695.
3. Marr, D. (1982). Vision (Freeman).
4. Qiu, F.T., Sugihara, T., and von der Heydt, R. (2007). Figure-ground mechanisms provide structure for selective attention. Nat. Neurosci. *10*, 1492–1499.
5. Martin, A.B., and von der Heydt, R. (2015). Spike synchrony reveals emergence of proto-objects in visual cortex. J. Neurosci. *35*, 6860–6870.
6. Lamme, V.A. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. J. Neurosci. *15*, 1605–1615.
7. Poort, J., Self, M.W., van Vugt, B., Malkki, H., and Roelfsema, P.R. (2016). Texture segregation causes early figure enhancement and later ground suppression in areas V1 and V4 of visual cortex. Cereb. Cortex *26*, 3964–3976.
8. Zipser, K., Lamme, V.A., and Schiller, P.H. (1996). Contextual modulation in primary visual cortex. J. Neurosci. *16*, 7376–7389.

9. Self, M.W., van Kerkoerle, T., Supèr, H., and Roelfsema, P.R. (2013). Distinct roles of the cortical layers of area V1 in figure-ground segregation. Curr. Biol. *23*, 2121–2129.
10. Supèr, H., Spekreijse, H., and Lamme, V.A. (2001). Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). Nat. Neurosci. *4*, 304–310.
11. Self, M.W., Kooijmans, R.N., Supèr, H., Lamme, V.A., and Roelfsema, P.R. (2012). Different glutamate receptors convey feedforward and recurrent processing in macaque V1. Proc. Natl. Acad. Sci. USA *109*, 11031–11036.
12. Kim, S.-H., and Feldman, J. (2009). Globally inconsistent figure/ground relations induced by a negative part. J. Vis. *9*, 1–13.
13. Roelfsema, P.R., Lamme, V.A., Spekreijse, H., and Bosch, H. (2002). Figure-ground segregation in a recurrent network architecture. J. Cogn. Neurosci. *14*, 525–537.
14. Self, M.W., Mookhoek, A., Tjalma, N., and Roelfsema, P.R. (2015). Contextual effects on perceived contrast: figure-ground assignment and orientation contrast. J. Vis. *15*, 2.
15. Rubin, E. (1921). Visuell wahrgenommene Figuren (Gyldendalske).
16. Coren, S. (1969). Brightness contrast as a function of figure-ground relations. J. Exp. Psychol. *80*, 517–524.
17. Poort, J., Raudies, F., Wannig, A., Lamme, V.A., Neumann, H., and Roelfsema, P.R. (2012). The role of attention in figure-ground segregation in areas V1 and V4 of the visual cortex. Neuron *75*, 143–156.
18. Lamme, V.A., Rodriguez-Rodriguez, V., and Spekreijse, H. (1999). Separate processing dynamics for texture elements, boundaries and surfaces in primary visual cortex of the macaque monkey. Cereb. Cortex *9*, 406–413.
19. Nothdurft, H.C., Gallant, J.L., and Van Essen, D.C. (2000). Response profiles to texture border patterns in area V1. Vis. Neurosci. *17*, 421–436.
20. Britten, K.H., Newsome, W.T., Shadlen, M.N., Celebrini, S., and Movshon, J.A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. Vis. Neurosci. *13*, 87–100.
21. Nienborg, H., and Cumming, B.G. (2009). Decision-related activity in sensory neurons reflects more than a neuron's causal effect. Nature *459*, 89–92.
22. Koffka, K. (1935). Principles of Gestalt Psychology (Psychology Press).
23. Wertheimer, M. (1938). Laws of Organization in Perceptual Forms (Routledge and Kegan Paul).
24. Roelfsema, P.R. (2006). Cortical algorithms for perceptual grouping. Annu. Rev. Neurosci. *29*, 203–227.
25. Corthout, E., and Supèr, H. (2004). Contextual modulation in V1: the Rossi-Zipser controversy. Exp. Brain Res. *156*, 118–123.
26. Rossi, A.F., Desimone, R., and Ungerleider, L.G. (2001). Contextual modulation in primary visual cortex of macaques. J. Neurosci. *21*, 1698–1709.
27. Jehee, J.F., Lamme, V.A., and Roelfsema, P.R. (2007). Boundary assignment in a recurrent network architecture. Vision Res. *47*, 1153–1165.
28. Kogo, N., and Wagemans, J. (2013). The "side" matters: how configurality is reflected in completion. Cogn. Neurosci. *4*, 31–45.
29. Qiu, F.T., and von der Heydt, R. (2007). Neural representation of transparent overlay. Nat. Neurosci. *10*, 283–284.
30. Zhou, H., Friedman, H.S., and von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. J. Neurosci. *20*, 6594–6611.
31. Li, W., Piëch, V., and Gilbert, C.D. (2008). Learning to link visual contours. Neuron *57*, 442–451.
32. Yan, Y., Rasch, M.J., Chen, M., Xiang, X., Huang, M., Wu, S., and Li, W. (2014). Perceptual training continuously refines neuronal population codes in primary visual cortex. Nat. Neurosci. *17*, 1380–1387.
33. Freedman, D.J., Riesenhuber, M., Poggio, T., and Miller, E.K. (2006). Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. Cereb. Cortex *16*, 1631–1644.
34. Kuai, S.-G., Levi, D., and Kourtzi, Z. (2013). Learning optimizes decision templates in the human visual cortex. Curr. Biol. *23*, 1799–1804.
35. Sáry, G., Vogels, R., and Orban, G.A. (1993). Cue-invariant shape selectivity of macaque inferior temporal neurons. Science *260*, 995–997.
36. Peterson, M.A., Harvey, E.M., and Weidenbacher, H.J. (1991). Shape recognition contributions to figure-ground reversal: which route counts? J. Exp. Psychol. Hum. Percept. Perform. *17*, 1075–1089.
37. Kourtzi, Z., Betts, L.R., Sarkheil, P., and Welchman, A.E. (2005). Distributed neural plasticity for shape learning in the human visual cortex. PLoS Biol. *3*, e204.
38. Nienborg, H., and Cumming, B. (2010). Correlations between the activity of sensory neurons and behavior: how much do they tell us about a neuron's causality? Curr. Opin. Neurobiol. *20*, 376–381.
39. Supèr, H., and Roelfsema, P.R. (2005). Chronic multiunit recordings in behaving animals: advantages and limitations. Prog. Brain Res. *147*, 263–282.
40. Cohen, M.R., and Maunsell, J.H. (2009). Attention improves performance primarily by reducing interneuronal correlations. Nat. Neurosci. *12*, 1594–1600.
41. Palmer, C., Cheng, S.-Y., and Seidemann, E. (2007). Linking neuronal and behavioral performance in a reaction-time visual detection task. J. Neurosci. *27*, 8122–8137.
42. Roelfsema, P.R., and Spekreijse, H. (2001). The representation of erroneously perceived stimuli in the primary visual cortex. Neuron *31*, 853–863.
43. Cavanaugh, J.R., Bair, W., and Movshon, J.A. (2002). Selectivity and spatial distribution of signals from the receptive field surround in macaque V1 neurons. J. Neurophysiol. *88*, 2547–2556.
44. Knierim, J.J., and van Essen, D.C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. J. Neurophysiol. *67*, 961–980.
45. Nelson, J.I., and Frost, B.J. (1978). Orientation-selective inhibition from beyond the classic visual receptive field. Brain Res. *139*, 359–365.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Organisms/Strains		
Macaca Mulatta	Biological Primate Research Center	https://bprc.nl/nl/home
Software and Algorithms		
MATLAB R2015a	Mathworks inc.	https://www.mathworks.com
Cogent Graphics	Vislab, UCL	http://www.vislab.ucl.ac.uk/cogent_graphics.php
Electrophysiological data	Mendeley Data	https://doi.org/10.17632/w9cvs4zwt5.1
Eye-movement data	Mendeley Data	https://doi.org/10.17632/w9cvs4zwt5.1
Behavioral data	Mendeley Data	https://doi.org/10.17632/w9cvs4zwt5.1
Custom MATLAB scripts	Mendeley Data	https://doi.org/10.17632/w9cvs4zwt5.1
Tucker-Davis Technologies recording system	Tucker-Davis Technologies	RRID:SCR_006495
Infrared video eye tracker	Thomas Recoding GmbH	ET49 https://www.thomasrecording.com/products/neuroscience-products/eye-tracking-systems/et-49-500hz.html

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Dr. Matthew Self (m.self@nin.knaw.nl).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All procedures complied with the NIH Guide for Care and Use of Laboratory Animals, and were approved by the institutional animal care and use committee of the Royal Netherlands Academy of Arts and Sciences. Two male macaque monkeys participated in the experiment. Monkeys were socially housed in pairs. One animal (monkey D) was completely naive and had only been trained on basic fixation tasks with luminance defined stimuli. The other (monkey J) had previously been trained for a figure-ground study with square figures and had performed a curve-tracing task [17]. The monkeys were socially housed in stable pairs in a specialized primate facility with natural daylight, controlled humidity and temperature. The home-cage was a large floor-to-ceiling cage which allowed natural climbing and swinging behavior. The cage had a solid floor, covered with sawdust and was enriched with toys and foraging items. Their diet consisted of monkey chow supplemented with fresh fruit. Their access to fluid was controlled, according to a carefully designed regime for fluid uptake. During weekdays the animals received water or dilute fruit juice in the experimental set-up upon correctly performed trials. We ensured that the animals drank sufficient fluid in the set-up and supplemented the animals with extra fluid after the recording session if they failed to drink enough. In the week-end the animals received at least 700ml of water in the home-cage supplied in a drinking bottle. The animals were regularly checked by veterinary staff and animal caretakers and their weight and general appearance were recorded in an electronic logbook on a daily basis during fluid-control periods.

METHOD DETAILS

Training history and surgical details

We implanted both monkeys with a titanium head-post (Crist instruments) under aseptic conditions and general anesthesia as reported previously [11]. The monkeys were trained to fixate on a 0.5° diameter fixation dot and hold their eyes within a small fixation window (1.1° diameter). They then underwent a second operation to implant 4x4 (2 in monkey D, 3 in monkey J), 4x5 (3 in monkey D) and 5x5 (2 in monkey J) arrays of micro-electrodes (Blackrock Microsystems) over opercular V1 (in total, 5 V1 arrays in each monkey) [17]. Four of the V1 arrays provided functional data in monkey D yielding a total of 72 channels and two of the V1 arrays were functional in monkey J, yielding 40 channels. Each monkey also had three arrays implanted in V4, in monkey D these were non-functional, in monkey J we recorded data from these arrays which will be reported in a later paper. The inter-electrode spacing on the arrays was 400μm. We first obtained pre-training data using a passive fixation task. The animals were then trained to perform the n/u discrimination task. After training we first obtained the shape-discrimination task data, followed by the post-training, passive fixation data and finally the attention-control data. The total number of recording days for each experimental phase and the range between the first and last recording day are provided in Table S1. We recorded activity from 2 arrays in monkey J and from 4 arrays in monkey D.

Electrophysiology

We recorded the envelope of multi-unit activity by digitizing the raw signal referenced to a subdural electrode at 24.4kHz. The raw signal was then band-pass filtered (2nd order Butterworth filter, 500Hz-5KHz) to isolate high-frequency (spiking) activity. This signal was rectified (negative becomes positive) and low-pass filtered (corner frequency = 200Hz) to produce the envelope of the high-frequency activity, which we refer to as MUA. The MUA signal reflects the population spiking of neurons within 100-150μm of the electrode and the population responses are very similar to those obtained by pooling across single units [39–41].

Stimuli

Stimuli were presented on a CRT monitor at a refresh rate of 85Hz and with a resolution of 1024x768 pixels viewed from a distance of 52cm (monkey D) or 75cm (monkey J). The monitor had a width of 40cm, yielding a field-of-view of 41.6° x 31.2° (monkey D) or 29.8° x 22.4° (monkey J). All stimuli were created using the COGENT graphics toolbox (developed by John Romaya at the LON at the Wellcome Department of Imaging Neuroscience) running in MATLAB (Mathworks Inc.). At the start of the trial the screen was gray (21.2 $cd.m^{-2}$) with a red fixation point. After 300ms of fixation the stimulus appeared. The stimuli consisted of full-screen textures composed of 13,000 oriented white lines (1.3° in length) drawn on a black background. Each day eight randomly generated textures were generated, four with an orientation of 45° and four with an orientation of 135°. To generate the n/u figures, a portion of one texture was copied over a background texture of the opposite orientation. We chose the textures to generate the figure and background pseudo-randomly so that on average each texture was presented an equal number of times, ensuring that on average the same texture elements were present at every location for figure and ground locations (with the exception of boundaries where the two orientations abut). The n/u form consisted of two legs of 8° in height and 4° in width, connected with a segment with a height of 2°. The central region between the legs was 4° in width, making the total figure size 12° x 8°. The connecting segment could be positioned at the top of the legs to create an n or at the bottom to create a u, we therefore refer to it as the discriminant segment. We also created a checkerboard texture which consisted of checks of 4° (width) by 6° (height) with alternating orientations. Across trials, we shifted the position of the n/u forms and the checkerboards horizontally in steps of 2° so that the RFs fell on different parts of the stimulus (thus creating 5 main conditions; Figures 2D and S6A). To study responses to the discriminant segment, we shifted the n/u forms vertically (only in monkey J). In monkey J, the n and u forms spatially overlapped so that the monkey could not use positional cues to determine his choice. In monkey D, the V1 RFs were more widely spaced and the n and u forms were shifted vertically to enclose these positions (Figure S5A). We ensured that the monkeys could generalize their performance across many vertical positions before recording the post-training neural data.

In the passive fixation task, the monkeys viewed the textures and maintained fixation for the duration of the trial (400ms). In the active version of the task the animals had to discriminate between texture-defined n and u forms by making an eye movement to one of two magenta targets positioned at the vertical meridian at 8° eccentricity. The monkeys viewed the textures for 400ms at which point the fixation point was removed (note that in monkey D the fixation condition was not fully enforced due to a programming error, see below). Monkey J then had to make an upward saccade for n's and a downward saccade for u's, in monkey D the response mapping was reversed. On trials in which a checkerboard was presented, the animals were rewarded for maintaining fixation for a further 275ms after the fixation dot was extinguished. The total number of trials performed for each condition and phase of the experiment are shown in Data S1.

Receptive Field Mapping

We mapped the RFs of each multi-unit site in V1 using a drifting luminance-defined bar that moved in one of four directions. The response to each direction was fitted with a Gaussian function. The borders of the RF were then calculated as described previously [39]. The signal-to-noise ratio (SNR_{RF}) of the response was then taken as the peak of the Gaussian divided by the standard deviation of the pre-trial baseline response. Only RFs for which the responses to all four bar directions had an SNR_{RF} of over 1 were included in the analyses. The median V1 RF size, taken as the square-root of the area, was 1.4° (range 1.0° to 3.1°) and the median eccentricity of the RFs was 4.3° (range = 2.9° to 6.1°).

QUANTIFICATION AND STATISTICAL ANALYSIS

Data Analysis

The MUA data from each recording site was normalized on each recording day. We first subtracted the mean activity in each pre-trial period in which the animal was fixating (−200 to 0ms relative to stimulus onset) to remove the baseline activity. We then normalized all responses to the maximum smoothed (26ms Gaussian kernel) peak response (in the time period 30-80ms after stimulus onset), averaging responses across all texture conditions. The data are therefore expressed in normalized units, i.e., a value of 0.1 indicates 10% of the difference in MUA between the peak and the baseline. Occasionally electrodes would show poor or noisy signals on individual days, to ensure these days did not affect the grand average we measured the signal-to-noise ratio (SNR_{DAY}) of the recording on each day. SNR_{DAY} was estimated by dividing the maximum of the initial peak response (30-80ms, averaged across all trials of a single day, smoothed with a 20 sample sliding window) by the standard deviation of the baseline activity across trials. Recording days on which $SNR_{DAY} < 1$ were removed from the analysis. Otherwise, responses from the same electrode were reliable across days as judged by similar SNR_{DAY} values and the general shape of the response on each day. We therefore averaged the remaining normalized responses across days.

We also calculated the signal-to-noise ratio for each site over days ($SNR_{ALLDAYS}$). Only recording sites with an $SNR_{ALLDAYS} > 4$ and a well-defined RF (see above) were included in the analysis. After applying these inclusion criteria, we analyzed data from 56 (out of 72) recording sites in V1 in monkey D and 38 sites (out of 42) in V1 in monkey J. For each recording site, we calculated the mean response across recording days for all trials belonging to each of the main conditions in the experiment, but we only included conditions in which the figure was positioned appropriately relative to the RF (Figure S2).

For statistical analyses, we took the mean activity of each site in two time windows: a peak window (30-80ms) and a sustained window (100-250ms). Note that while the animal was required to retain fixation for 400ms, monkey D was able to make eye-movements after 200ms due to a technical error and we therefore only analyzed data from 0-250ms (allowing for the visual latency of responses in V1, and we also removed the very few trials on which the saccade was made earlier than 210ms (0.5% of all trials) to ensure the absence of eye-movement related activity), and we used the same analysis window for the data of monkey J so that we could compare and pool the data across monkeys (the results remained the same if the entire 400ms time period was used for monkey J). Statistics were calculated at two different levels in the manuscript. For individual channels (e.g., the example channel in Figure 2A) the statistics were performed across trials. For the remainder of the manuscript the statistics were performed across recording channels. In all cases the normality of the sampling distribution was checked using the Shapiro-Wilk test before applying parametric statistics. For the analyses of the effect of spatial position across sites we used repeated-measures ANOVAs with the 5 RF positions as a factor. We compared stimulus conditions using post hoc t tests, corrected for multiple comparisons using the Bonferroni method. We compared different epochs (before and after training) using a two-way repeated-measures ANOVAs with the factors RF position (5 levels) and epoch (2 levels).

To compute the spatio-temporal profile of neuronal activity we first calculated the horizontal distances between the RF-center and the center of the proto-figure and proto-ground of the n/u shape or the individual checks of the checkerboard stimulus, binned into 0.75° wide bins (Figures S5A–S5D). We extracted the MUA time-course at each location and subtracted the mean response in the background condition (mean of all trials with the figure center more than 8° away from the RF center). We then averaged activity elicited by n's and u's across the sites within each bin and linearly interpolated the resulting response on a spatio-temporal grid (spatial resolution: 0.25°, temporal resolution: 2ms). We only included conditions for which the vertical distance between figure center and RF center was less than 1.5°, in order to exclude responses elicited by the internal boundary of the discriminant segment (more details have been specified in Figures S2 and S5). For statistical analysis of the spatio-temporal maps (Figure 5C, Figure 6E) we regrouped the interpolated-data into the five main conditions according to the x-position of the RF. Data from −9° to −7° and 7° to 9° were assigned to the ground, from −7° to −5° and 5° to 7° to the outer edge, from −5° to −3° and 3° to 5° to the proto-figure, from −3° to −1° and 1° to 3° to the inner edge and from −1° to 1° to the proto-ground.

The latency of POM was calculated by conducting paired t tests between the [proto-figure - check] and [proto-ground - check] subtractions. One test was performed per sample of the unsmoothed time-series. We took the latency as the first sample with a significant result ($p < 0.05$, uncorrected) that was followed by a further 10 significant samples. The latency of the coarse proto-object enhancement (red/blue lines in Figure 3D) was calculated using the same approach, except that a one-sample t test was used (e.g., proto-figure – check versus 0). Statistical tests between latencies were performed by resampling the recording sites 1000 times with replacement and performing the same procedure as described above to determine the latencies of each bootstrap sample for the different conditions. The p value of the latency difference between two conditions was estimated by taking the difference in latency for each bootstrap sample and taking the fraction of samples on which the difference was less than or equal to zero (one-tailed test).

To calculate choice-probabilities for each recording site we calculated the area under the receiver operating characteristic curve based on V1 activity in a time window from 100-250ms, comparing trials in which the monkey gave a correct response or made an error, as described previously [20, 42]. Activity from different recording days was first normalized as described above before data across all recording days per site was entered into the analysis. According to the definition used by us, the choice probability is higher than 0.5 if correct trials are associated with higher V1 activity, and smaller if correct trials are associated with weaker V1 activity. A choice probability of 0.5 implies that V1 activity was uninformative about accuracy.

Analysis of eye position

Eye movements were recorded using a digital-camera (Thomas recordings, 250Hz frame-rate). For analysis of the eye data the position of the pupil was digitized and recorded at 500Hz by the TDT recording system, the eye position traces were low-pass filtered at 100Hz for the analysis of eye position. We examined whether the difference in V1 response elicited by the proto-figure and proto-ground condition was caused by systematic variations in gaze position within the fixation window. We calculated for each array the mean eye position in trials with the proto-figure or proto-ground centered on the RF in every trial (time-window: 0-250ms after stimulus onset) (Figure S4). We did not observe significant differences in eye x or y position in the experiments for any of the arrays for either monkey (two-sample t test, all $p > 0.1$). Micro-saccades were detected using a velocity threshold. The x and y eye traces were first converted into velocity by taking the difference between subsequent samples. This velocity measure was smoothed with a 20ms sliding window and then converted into degrees-of-visual-angle per second. If the maximum eye velocity (in either x or y) in the time window (0-250ms) was greater than 50 $deg.s^{-1}$ a micro-saccade was said to have occurred and the trial was removed from the analysis. This occurred on 0.6% of trials for monkey D and 0.9% of trials for monkey J.

Analysis of feature contrast

It is well-known that cells in V1 are suppressed by stimuli in their surround. This suppression is strongest when image elements in the surround of the neurons' RF have the same orientation as those in the center [43–45]. Large regions with a uniform orientation (such as background regions) therefore elicit greater suppression than small regions with feature contrast (such as figural regions). Previous experiments examining figure-ground modulation with square figures did not dissociate figure-ground modulation from the influence of feature-contrast.

We examined whether feature contrast could be used to assign figure and background with the N/U stimuli with a matrix of 500x500 pixels with a spacing of 0.1°, centering the RF on the proto-ground, the proto-figure or a 4° square (for comparison with earlier studies). Each pixel either had the same orientation as the center of the RF (coded as a 1) or the orthogonal orientation (0). We then measured feature contrast for a series of circular apertures with diameter d ranging from 0.4 to 40° in steps of 0.2°. We calculated feature-contrast FC as:

$$FC(d) = \frac{\sum_{p \in C_d} (1 - F_p)}{n_{p \in C_d}}$$

Where Fp is the feature at pixel p (with a value of 1 if it matches the feature of the RF and 0 otherwise) that fell within the circular Cd aperture of diameter d and $n_{p \in C_d}$ was the total number of pixels inside the aperture. FC is zero when all pixels match the orientation at the center of the RF and 1 when all pixels have the orthogonal orientation. If the RF falls on the background, the FC is usually close to zero, it has a value close to 0.5 at a straight boundary between figure and ground and a value near one in a small region that pops out due to the feature contrast.

We found that FC was higher for the proto-ground than for the proto-figure for apertures smaller than 11.1° (Figure S1). For these smaller apertures, feature-contrast would therefore bias the proto-ground to be incorrectly designated as a figure. The visual system would have to integrate over regions of greater than 11.1° in diameter to correctly assign image elements of the N/U stimuli to figure and ground based on feature-contrast.

DATA AND SOFTWARE AVAILABILITY

The electrophysiological, behavioral and eye-movement data reported in this manuscript and custom MATLAB analysis scripts will be available from Mendeley Data after publication at the following https://doi.org/10.17632/w9cvs4zwt5.1