The representation of occluded image regions in area V1 of monkeys and humans

Papale, Paolo; Wang, Feng; Morgan, A Tyler; Chen, Xing; Gilhuis, Amparo; Petro, Lucy S; Muckli, Lars; Roelfsema, Pieter R; Self, Matthew W

published in
Current Biology
2023

DOI (link to publisher)
10.1016/j.cub.2023.08.010

document version
Early version, also known as pre-print

The representation of occluded image regions in area V1 of monkeys and humans

Paolo Papale[1,‡,*], Feng Wang[1,‡], A. Tyler Morgan[2,3,5], Xing Chen[4], Amparo Gilhuis[1], Lucy S. Petro[2,5], Lars Muckli[2,5], Pieter R. Roelfsema[1,6,7,8,¶,*,§] and Matthew W. Self[1,¶]

1. Department of Vision & Cognition, Netherlands Institute for Neuroscience (KNAW), 1105 BA Amsterdam, Netherlands.
2. Centre for Cognitive NeuroImaging, School of Psychology and Neuroscience, College of Medical, Veterinary and Life Sciences, University of Glasgow, 62 Hillhead Street, Glasgow, G12 8QB, UK.
3. Current address: Section on Functional Imaging Methods, Laboratory of Brain and Cognition, National Institute of Mental Health, Bethesda, Maryland, USA.
4. Department of Ophthalmology, University of Pittsburgh School of Medicine, 203 Lothrop St, PA 15213, Pittsburgh, US
5. Imaging Centre for Excellence (ICE), College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G51 4LB, UK.
6. Department of Integrative Neurophysiology, VU University, De Boelelaan 1085, 1081 HV Amsterdam, Netherlands.
7. Department of Psychiatry, Academic Medical Centre, Postbus 22660, 1100 DD Amsterdam, Netherlands.
8. Laboratory of Visual Brain Therapy, Sorbonne Université, INSERM, CNRS, Institut de la Vision, 17 rue Moreau, F-75012 Paris, France,

‡ shared first author contribution
¶ shared senior author contribution
§ Lead Contact
* correspondence should be addressed to: *p.roelfsema@nin.knaw.nl*, *p.papale@nin.knaw.nl*
Twitter: @Pieters_Tweet, @paolo_papale

Summary

Neuronal activity in the primary visual cortex (V1) is driven by feedforward input from within the neurons' receptive fields (RFs) and modulated by contextual information in regions surrounding the RF. The effect of contextual information on spiking activity occurs rapidly and is therefore challenging to dissociate from feedforward input. To address this challenge, we recorded the spiking activity of V1 neurons in monkeys viewing either natural scenes or scenes where the information in the RF was occluded, effectively removing the feedforward input. We found that V1 neurons responded rapidly and selectively to occluded scenes. V1 responses elicited by occluded stimuli could be used to decode individual scenes and could be predicted from those elicited by non-occluded images, indicating that there is overlap between visually driven and contextual responses. We used representational similarity analysis to show that the structure of V1 representations of occluded scenes measured with electrophysiology in monkeys correlates strongly with the representations of the same scenes in humans measured with fMRI. Our results reveal that contextual influences rapidly alter V1 spiking activity in monkeys over distances of several degrees in the visual field, carry information about individual scenes and resemble those in human V1.

Introduction

Neurons in primary visual cortex (V1) respond to specific features that are present in a small portion of visual space, known as their receptive field (RF). The feature and spatial selectivity of the RF depends on the feedforward inputs that the V1 neuron receives from the retina, via the lateral geniculate nucleus of the thalamus. The initial responses of cells in V1 are therefore largely determined by the characteristics of the stimulus within their RF. Recent studies demonstrated that the tuning of V1 neurons can be characterized well by taking advantage of developments in artificial intelligence, in particular convolutional feedforward networks[1]. However, V1 cells also receive numerous lateral and feedback projections that provide contextual input from the rest of the visual field and that are lacking from purely feedforward models[2,3]. Accordingly, the feedforward response of V1 neurons is rapidly modulated by these contextual inputs, and it is challenging to dissociate the influence of feedforward and contextual inputs[4]. Recent studies have used partially occluded images to reveal the influence of visual context on V1 representations in the absence of information in the RF[5–7]. If the occluder is placed over the neurons' RF, their feedforward input is removed but the contextual inputs remain[5].

Several neuroimaging studies have presented partially occluded images to humans, showing that V1 hemodynamic activity contains information about individual stimuli, even in the absence of feedforward information[5,8–10]. However, fMRI offers limited spatial and temporal resolution and the precise relationship between hemodynamic activity and spiking activity is not fully understood[11]. It therefore remains unknown whether the spiking activity of V1 neurons encodes contextual input in the absence of feedforward drive and how these contextual influences unfold across time.

Results

To examine the representation of occluded images in V1 spiking activity, we presented 24 images, depicting natural (beaches, forests and mountains) and man-made scenes (buildings, highways and industry), which had been used in previous neuroimaging studies[8], to two awake fixating macaque monkeys. The monkeys had not seen these images prior to the experiments. The images were either partially occluded ('occluded' condition) or not ('non-occluded' condition; Figure 1A,B). We recorded multi-unit spiking activity (MUA) from a total of 175 recording sites in V1 with RFs located on the occluded region of the images and measured responses to non-occluded and occluded images.

Different V1 responses in occluded image regions

We first examined the mean V1 activity across recording sites elicited by each of the 24 scenes (Figure 1C). As expected, the different images elicited distinct activity levels in the 'non-occluded' condition because the information inside the neurons' RFs varied. Neurons exhibited weaker responses in the occluded condition ($p < 0.001$, Wilcoxon signed rank test), which were driven by the increase in luminance in the RF when the white occluder replaced the gray screen in the RF before the stimulus appeared (Figure 1B). The latency of the response to the occluder was 44ms and it was not different from the latency of the response evoked by non-occluded stimuli (41ms, $p = 0.28$, Wilcoxon signed rank test). The occluder in the neurons' RF was identical for all stimuli and differences in activity elicited by different stimuli arose later for occluded images, presumably due to extra delays associated with recurrent connections[2,3,12,13]. We therefore also evaluated the latency of stimulus selectivity, i.e. the moment at which differences in responses elicited by the pictures first emerged. To

this aim we determined the average response elicited by each of the 24 scenes at ~~for~~ each recording site and computed the 276 pairwise differences (between scene 1 and scene 2, scene 1 and scene 3, etc.). We determined the latency of each difference response~~s~~ with a curve fitting method and took the average as the recording site's response latency (Figure S1A,B). An advantage of this approach is that the curve fitting method decouples the latency from response strength (signal-to-noise ratio; see Methods and Figure S1). In the non-occluded condition, the median latency of stimulus selectivity across recording sites was 53ms, which was shorter than the latency of selectivity of 64ms for the occluded condition ($p < 0.001$, Wilcoxon signed rank test). The latency of stimulus selectivity for the non-occluded stimulus is in accordance with previous results[13], and the latency of 64ms of the contextual influence is in the same range as the latency of tuning to border-ownership in areas V1 and V2[14,15], but longer than the latency of contextual effects in figure-ground paradigms (~100ms, e.g. ref. [16]). The contextual effect could be caused by horizontal connections from V1 neurons with RFs on the non-occluded image regions or by feedback from higher visual areas. The main prediction of models based on horizonal propagation is that the latency of the contextual effect increases with the distance between a neuron's RF and the border of the occluder, because of the limited conduction velocity of horizontal connections[17–23]. However, the latency in the occluded condition did not depend on the distance between the RF and border of the occluder (Figure 1D; linear regression: $R^2 = 0.001$; $p = 0.64$, F-test), suggesting that the contextual response arrives simultaneously across the V1-representation of the occluded region. Such a pattern of latencies is consistent with a feedback effect from downstream visual areas, as feedback effects can span large areas of the visual scene and are therefore relatively independent of the distance between the RF and the border of the occluder[17].

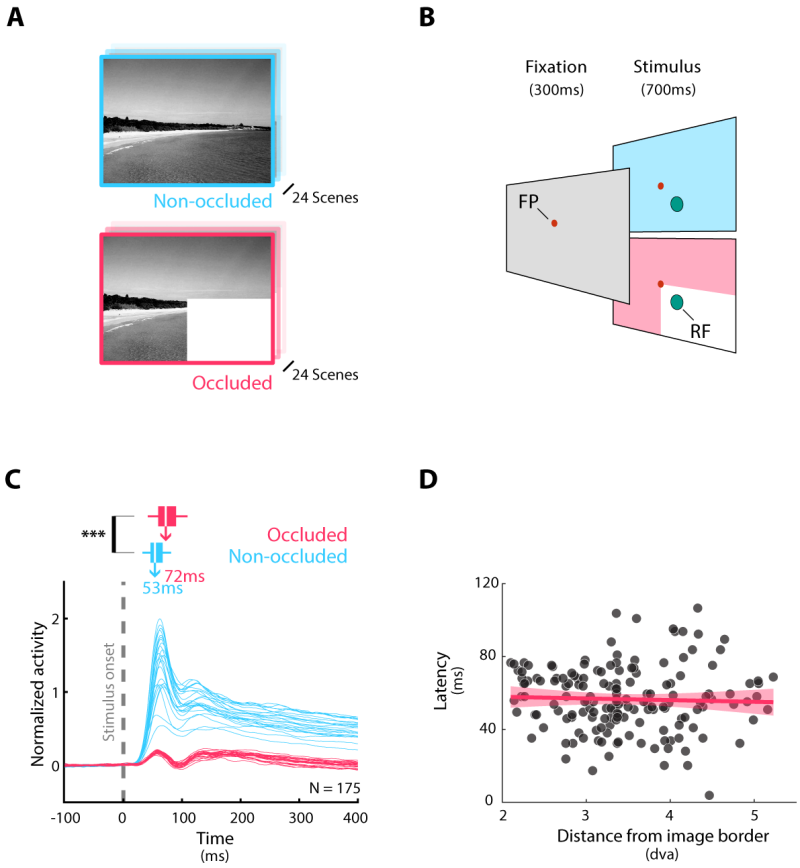

Figure 1. V1 response to non-occluded and partially occluded natural scenes.

(A) We presented 24 natural scenes to the monkeys; either non-occluded (top, blue) or partially occluded by a white rectangle placed in the bottom-right quadrant (bottom, pink). (B) We recorded V1 spiking activity from two monkeys. After they maintained their gaze on a red dot (on a gray background) for 300ms, either a non-occluded (top, blue) or partially occluded (bottom, pink) stimulus was presented for 700ms. (C) Mean response elicited by each of the 24 scenes (blue: non-occluded; pink: occluded condition). Traces represent the V1 population responses elicited by each scene. We computed the latency of the stimulus selectivity based on difference responses, which was longer for the occluded than the non-occluded conditions (***: $p < 0.001$, Wilcoxon signed rank test); bars represent the distribution of stimulus selectivity latencies across recording sites (white line = median, box = inter-quartile range, whiskers = min/max of data). (D) We examined the influence of the distance between the RFs and the non-occluded portion of the image on the latency of the stimulus selectivity of individual recording sites in the occluded condition with a linear regression. This relationship was not significant (N= 175 recording sites; $R^2 = 0.001$; $p = 0.64$, F-test; red line, fit; red-shaded region, 95%-confidence bounds; see also: Figure S1).

Decoding analysis

We next examined how well V1 spiking activity discriminated between the 24 scenes using a multi-class linear classifier. We trained independent classifiers at each time-point from 100ms before stimulus onset until 400ms thereafter (the data was first smoothed with Gaussian kernel with an s.d. of 25ms). The classifier was trained and cross-validated ~~on~~ using independent trials (~200 per stimulus, ~160 for training and ~40 for cross-validation) and we repeated the procedure both within and across the two stimulus conditions (Figure 2). As

expected, training and testing on the non-occluded images yielded a high accuracy (Figure 2A, blue; $p<0.05$, permutation test, false discovery rate (FDR) corrected) that was stable over time. In the occluded condition (Figure 2A, pink) decoding of the images in V1 was significant and stable over time although the accuracy was lower than for the non-occluded images ($p < 0.001$, Wilcoxon signed rank test)[4,7]. We next examined if the influence of the occluded stimulus on the V1 responses depended on the distance between the RF and the border of the occluder (Figure S2E), assessing the decoding accuracy at individual recording sites. Interestingly, the decoding accuracy decreased with the distance between the RF and the border (linear regression: $R^2 = 0.08$; $p < 0.01$, F-test), indicating that neurons with RFs that were closer to non-occluded regions conveyed more information about the scene than neurons with RFs at a larger distance.

To examine whether the V1 representation of occluded images resembles that of non-occluded images, we examined the cross-decoding performance. We trained a decoder on the data from the non-occluded images and tested it on data from the occluded images, and vice versa. We observed significant generalization in both directions (Figure 2B; $p<0.05$, permutation test, FDR corrected), although the accuracy was lower than for decoding within conditions. The median accuracy was higher when training on the occluded and testing on the non-occluded condition than vice versa ($p < 0.001$, Wilcoxon signed rank test across time). To explore whether the code was stable or variable, we examined the temporal generality of the decoders by repeating the procedure for each pair of time-points[18] (e.g. training at 100ms after stimulus onset and testing at 200ms, etc.). This additional analysis did reveal relatively stable representations (Figure S2). One difference between the conditions was that there was a change in luminance in the RF when the stimulus appeared in the occluded condition. In an additional experiment, we therefore used either a grey or white occluder and found that decoding was preserved when the occluder's luminance did not changed (Figure S3).

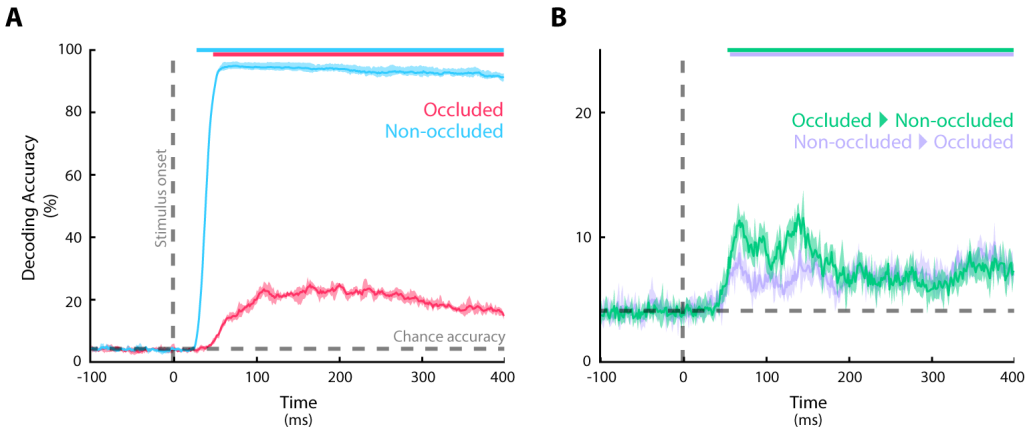

Figure 2. Neural decoding reveals similar contextual representations during V1 processing of non-occluded and occluded natural scenes.

(A) Decoding accuracy for the non-occluded (blue) and occluded (pink) pictures. (B) Cross-decoding results. In purple: non-occluded$_{train}$/occluded$_{test}$. In green: occluded$_{train}$/non-occluded$_{test}$. Colored bars at the top depict significant time-points ($p<0.05$, permutation test, false discovery rate corrected). Shaded regions show the minimum and maximum of repetitions of the cross-validation procedure. Horizontal dashed line, chance level. Vertical dashed line, stimulus onset. See also: Figure S2 and S3.

Comparison between monkeys and humans

Previous studies demonstrated that it is possible to use fMRI for the decoding of individual scenes based on the representation of occluded image regions in area V1 of humans[4,5,8]. Is there a relationship between the representations of occluded image regions measured with fMRI in humans and with electrophysiology in monkeys? To investigate this question, we compared the spiking activity recorded in the two monkeys to neuroimaging data obtained from 18 human subjects who viewed the same natural scenes that were used here[8] with a representational similarity analysis (RSA)[19,20]. For the human data, we computed dissimilarity matrices using the multi-voxel fMRI patterns in parts of human V1 that represented occluded portions of the scene (Figure 3B). Each entry in the representational dissimilarity matrix (RDM) of Figure 3B represents the dissimilarity of the multivoxel response patterns between two pictures.

The monkey data had a higher temporal resolution than the human fMRI data and we could therefore compute monkey RDMs at successive time-points (Figure 3A)[19]. To compare the similarity between monkey and human representations of occluded image regions, we calculated the Spearman's rank correlation between the monkey RDMs at successive time points and the human RDM (Figure 3C). We found a robust correlation that reached the noise ceiling ~100ms after stimulus onset, indicating that the correlation between V1 representations of monkeys and humans was as high as that between individual human participants and their average. These results demonstrate a remarkable similarity of V1 response patterns elicited by contextual information outside the neurons' RFs across species, using measurements from different modalities (fMRI responses in humans vs. spiking activity in monkeys).

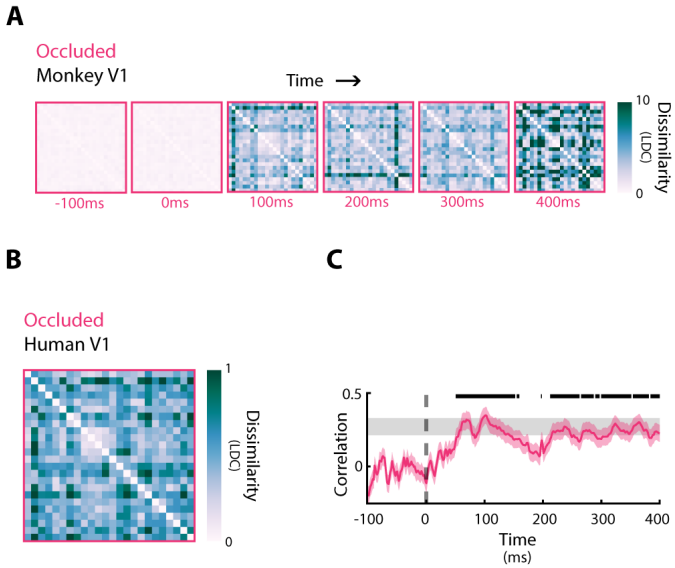

Figure 3. V1 representations in monkey and humans are correlated.

(A) RDM of monkey V1 spiking activity across time points. The 24 stimuli appear on the x- and y-axes of the RDM and the color of cells denote the linear discriminant contrast (LDC, see Methods), a measure of the similarity of activity patterns across recording sites elicited by two stimuli. (B) Average RDM of V1 voxels across 18 subjects in the occluded region of the same images that the monkeys saw, measured with fMRI. (C) The Spearman's correlation between monkey RDMs at successive time points and the human RDM. The black markers at top denote

time points with significant correlation ($p < 0.05$, permutation test, FDR corrected); the gray shaded region indicates the interval of the highest correlation that is theoretically possible given the noise in the data (i.e. noise ceiling); the shaded pink region represents bootstrapped SEM (1,000 iterations).

Discussion

Our results demonstrate contextual influences on spiking activity of V1 neurons that represent occluded image regions. The influence of visual scene information starts around 64ms after stimulus onset, which is in the same range as studies on border ownership[14,15] showing that the activity of neurons in early visual cortical areas that is elicited by an edge depends on the image region to which this edge belongs. At the same time, the latency of the occlusion responses is earlier than what has been observed in studies on contextual influences in V1 in texture-segregation tasks (e.g. refs.[16,21,22]). The latency of the contextual influences did not depend on the distance between the RF and the boundary of the occluder, which indicates that the scene information is brought to the neurons' RFs mostly by top-down rather than horizontal connections alone. Horizontal connections are shorter than the cortical distances over which the contextual effects were expressed. At an eccentricity of 5 dva, which was the approximate position of the RFs, the cortical magnification factor is ~2.5 mm/dva[23]. Hence, 5 dva corresponds to 12,5 mm of cortical distance, which is larger than the length of horizontal connections[24,25] so that the contextual effects could only be produced by propagation across several horizontal connections. Estimates of the propagation speed of horizontal connections in V1 vary between 5 and 25cm/s, as measured with intracellular recordings[26], in the local field potential[27], with voltage-sensitive dye imaging[28,29] and in spiking activity[30,31]. Hence, the predicted latency increase is between 10ms/dva and 50ms/dva at the eccentricities tested by us, which should have shown up in our analysis (Figure 1D).

Although information about the occluded scene arrived simultaneously across the V1 representation, neurons with RFs near the occluder represented more scene information than neurons with RFs that were farther from it, in accordance a previous human fMRI study[8]. The increase in RF sizes at higher cortical areas that send direct or indirect feedback to V1 may provide an explanation. V1 neurons with RFs close to the occluder may receive direct feedback from V2 and V3 neurons with RFs straddling the occluder's border, which is informative about image regions that are near the V1 RF. In contrast, V1 neurons with RFs far from the border may only receive feedback from higher areas with very large receptive fields, tuned to more abstract properties of the scene.

What is the structure of the information about individual scenes that is sent back to the occluded V1 representation? Studies in humans have suggested that this top-down signal correlates with a sketch of the visual features that might be present in the occluded region[8]. Studies in mice using occluded grating-stimuli have found evidence that feature-specific feedback signals alter spiking activity in V1[7,32]. Specifically, these studies positioned a circular occluder over a horizontal or vertical grating and demonstrated that it is possible to decode grating orientation by recording from V1 neurons with a RF at the occluder. Our cross-decoding results demonstrate overlap between the information conveyed by neurons with RFs on the occluded and non-occluded image regions, but with an asymmetry: decoders that were trained on non-occluded image regions did not generalize as well as decoders that were trained on occluded image regions. One possible explanation is that the signal that is present in occluded regions of V1 is dominated by semantic information about the visual scene and

about the classes of objects that are near the occluder. This information cannot predict the precise contents of the RF for non-occluded images. We note, however, that the differences between the two cross-decoders were relatively small and future studies will be needed to better understand the information content of the putative top-down signal. State-of the-art models for the tuning of V1 neurons are based on convolutional feedforward networks[1,33–35]. These models do not account for contextual influences, but future studies could try to extend these models with lateral and feedback connections, to explain contextual influences, including V1 tuning to partially occluded images.

Importantly, we found that V1 scene representations based on human neuroimaging data correlate strongly with spiking activity in monkeys, providing direct evidence of similar representations in the absence of feedforward information in the two species. This result represents an important step in bridging the complementary lines of research in humans and non-human primates. It suggests that what we know about the contextual modulation from experiments in monkeys is also likely to generalize to humans[36]. These results lay the groundwork for future research on the exact structure of the top-down signal that is sent back to occluded regions of V1, and on its potential role in shaping the activity in V1 regions devoid of feedforward information[37,38].

Acknowledgments

We thank Kor Brandsma, Anneke Ditewig, Taijsha van Rees and Lex Beekman for biotechnical support, and Rick Schuurman for assistance during surgeries. We thank Blackrock Microsystems for technical assistance and collaboration. The work was supported by NWO (STW-Perspectief grant P15-42 “NESTOR”, Crossover grant 17619 “INTENSE” and OCENW.XS22.2.097 and “DBI2”, a Gravitation program of the Dutch Ministry of Science), the European Union Horizon 2020 Framework Program under specific grant agreement 785907 and 945539 “Human Brain Project” SGA2 and SGA3, an ERC grant (101052963 “NUMEROUS”) and the Biotechnology and Biological Sciences Research Council (BBSRC) ‘Layer-specific cortical feedback’.

Author contribution

P.P., M.W.S and P.R.R conceived of the study. F.W., X.C., M.W.S., and A.G. performed the experiments. P.P. analysed data with contributions from F.W., A.T.M. and M.W.S. P.R.R. and L.M. (with L.S.P.) obtained funding. A.T.M., L.S.P. and L.M. provided data for the study. M.W.S and P.R.R oversaw all aspects of the project. P.P., P.R.R. and M.W.S wrote the paper with contributions from all authors.

Declaration of Interests

P.R.R. is founder and shareholder of Phosphoenix, a company that aims to develop a visual brain prosthesis for blind people.

STAR★Methods

Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Pieter Roelfsema (p.roelfsema@nin.knaw.nl).

Materials availability
This study did not generate new unique reagents.

Data and code availability
Data is available on EBRAINS at the link: https://doi.org/10.25493/KABE-GS0. Code is available on GitHub at the link: https://github.com/PPthe2nd/Occlusion_monkey.

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 (KNAW). Two macaque monkeys (males, 7 and 8 years old) participated in the experiments. They were socially housed in stable pairs in a specialized primate facility with natural daylight, controlled humidity and temperature. The home-cage was a large floor-to-ceiling cage which allowed natural climbing and swinging behavior. The cage had a solid floor, covered with sawdust and was enriched with toys and foraging items. Their diet consisted of monkey chow, supplemented with fresh fruit. Their access to fluid was controlled, according to a carefully designed regime for fluid uptake. During weekdays, the animals received water or diluted fruit juice in the experimental set-up upon correctly performed trials. We ensured that the animals drank sufficient fluid in the set-up and supplemented the animals with extra fluid after the recording session if they did not drink enough. During the weekend, they received a full bottle of water (700-940 ml per day) in the home cage. The animals were regularly checked by veterinary staff and animal caretakers and their weight and general appearance were recorded daily in an electronic logbook during fluid-control periods.

Method details
Surgical details
We implanted a 3D-printed titanium head-post on the monkey's skull under aseptic conditions and general anesthesia as reported previously[39]. The monkeys were trained to direct their gaze to a 0.2° diameter fixation dot and hold their eyes within a fixation window (1.1° diameter). They then underwent a second operation to implant arrays of micro-electrodes (Utah-probes, Blackrock Microsystems) over opercular V1 and V4 (eight 5x5 arrays in monkey B; sixteen 8x8 arrays in monkey L). For the present study we only used the recordings from the V1 arrays.

Electrophysiology
We recorded neuronal activity from 144 recording sites in V1 in monkey B (6 arrays) and 896 in monkey L (14 arrays). Neural signals were referenced to a subdural electrode and amplified using 32 channel Tucker-Davis Technologies ZIF-clip headstage amplifiers (monkey B) or Blackrock microsystems Cereplex-M headstage amplifiers. The amplified signal was sampled at 24.4kHz using a Tucker-Davis Technologies RZ2 system (monkey B) or 30kHz using a Blackrock microsystems system (monkey L). We measured the envelope of multi-unit activity by band-pass filtering the signal offline (2nd order Butterworth filter, 500 Hz-5 KHz, filtfilt.m in MATLAB) to isolate high-frequency (spiking) activity. This signal was rectified (negative becomes positive) and low-pass filtered (corner frequency = 200 Hz) to produce the envelope of the high-frequency activity, which we refer to as MUA[40]. The MUA signal was down-sampled to 770Hz and stored for further analysis. The MUA signal reflects the population

spiking of neurons within 100-150 µm of the electrode and the population responses are very similar to those obtained by pooling across single units[40,41].

Receptive Field Mapping

We mapped the RFs of each recording site in V1 using a drifting luminance-defined bar that moved in one of eight directions (every 45°). The response to each direction was fitted with a Gaussian function and the borders of the RFs were calculated as described previously[40].

Stimulus presentation

We used a dataset of 24 grayscale natural scenes and a version of the same images with the bottom-right quadrant occluded by a white rectangle (taken from ref. [8]). Images spanned several categories: beaches, buildings, forests, highways, industry, and mountains. Stimuli were presented in two different setups. For monkey B, we used a CRT monitor with refresh rate of 75 Hz and resolution of 1024x768 pixels, viewed from a distance of 49.5 cm. The monitor provided a field-of-view of 43.5 x 32.6°. For monkey L, we used a CRT monitor with refresh rate of 85 Hz and resolution of 1024x768 pixels viewed from a distance of 50 cm. The monitor provided a field-of-view of 39.6 x 29.7°. In both setups, the eye position was recorded using a digital camera (Thomas Recording, 250-Hz framerate) and stimuli were generated in Matlab using the COGENT Graphics toolbox (developed by John Romaya at the LON at the Wellcome Department of Imaging Neuroscience) and custom control software[42].

At the start of the trial, the monkey directed its gaze to a fixation point on a gray background (luminance 14 $cd \cdot m^{-2}$ for monkey B; 17.5 $cd \cdot m^{-2}$ for monkey L). We presented the image once the monkey had maintained fixation for 300ms, and the animal had to maintain fixation for an additional 700ms during stimulus presentation. Reward was delivered after every correct trial. Aborted trials (i.e., when the monkeys did not maintain fixation for 700ms) were repeated at the end of the recording session. Monkey were not exposed to the stimuli before the first recording session.

Selection of recording sites and inclusion of data

To normalize MUA, we first subtracted the mean activity obtained during the pre-trial period, during which the animal was fixating (-200 to 0ms relative to stimulus onset) and divided activity by the maximum smoothed (25ms Gaussian kernel) peak response (0-200ms from stimulus onset)[16,35]. We only included recording sites with a sufficient signal-to-noise ratio (SNR), estimated by dividing the maximum of the initial peak response by the standard deviation of the baseline activity across trials. If the SNR of a recording site was less than 1 across recording days, we removed that recording site from the analysis. In addition, we excluded recording sites whose RF overlapped with non-occluded parts of the stimuli in the occluded condition. We only included sites whose RFs were fully inside the occluder and whose RF centers were at least 2° away from both the horizontal and vertical meridian, so that a total of 175 recording sites remained (18 in monkey B and 157 in monkey L). The mean activity, response latency and decoding accuracy when training and testing on non-occluded stimuli were comparable across the two monkeys, thus we aggregated the data in the analyses (see Figure S3D for an analysis per monkey). Decoding was also significant in individual monkeys when training and testing on the occluded stimuli. When we used the average response after stimulus onset as input to the decoder, the accuracy for monkey B with 3 stimuli was 57% (N=18 recording sites; Figure S3; chance = 33%). The decoding accuracies were in the same range when we equated the number of recording sites between

the two monkeys (Figure S3). When we tested 24 stimuli and included all recording sites, the decoding accuracy was 8% (chance = 4.2%) for monkey B and 62% for monkey L. When we matched the number of recording sites by randomly selecting 18 V1 recording sites from the 157 in monkey L (N=100 iterations), the accuracy was 18% (all ps < 0.001, permutation test). In total, we recorded 9600 correct trials from monkey B (200 repetitions per stimulus) and 9,864 trials in monkey L (~205 repetitions per stimulus). For the main decoding analyses, data from both monkeys (and across days) was combined and the number of repetitions was balanced across stimuli by removing surplus repetitions at the end of the recordings.

Analyses of latency

To compute the latency of neural responses at individual recording sites, we computed the time-point at which the difference time-course between each pair of stimuli increased above zero. To this aim, we computed the difference in activity between every pair of stimuli in half of the trials (randomly selected) to determine which of the two stimuli gave the stronger response and fitted a function to the time-course of the difference in the other half of the trials[16,35,43]. The function was derived from the assumption that the onset of the response had a Gaussian distribution across trials, and that a fraction of the response dissipated exponentially, yielding the following equation:

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

where $G(\mathrm{t}, \mu, \sigma)$ is a cumulative Gaussian density with mean μ and standard deviation σ, α^{-1} is the time constant of the dissipation, and c and d represent the contribution of the non-dissipating and dissipating components. The function was fit using non-linear least squares (fit.m in MATLAB). The latency was defined as the point at which the fitted function reached 33% of its maximum.

To compute the pairwise differences in activity elicited by stimuli, we always subtracted the weaker response from the stronger response (averaged in the 0-400ms window) so that the difference was always positive (X-Y if X>Y, and Y-X if Y>X). The procedure to compute the latency is depicted in Figure S1. For every recording site, we computed the latency of each pairwise difference with the curve-fitting method (N=276) and then averaged the latencies. One advantage of the curve-fitting method is that is relatively insensitive to the SNR. Accordingly, the latency was not correlated with SNR across recording sites (R^2 = 0.018; p = 0.073, F-test).

Neural decoding

We used a neural decoding approach to investigate how well the responses discriminated between stimuli. We note that the result of such an analysis may differ from how regions within the primate brain read out of V1 activity to generate behavioral reports. All trials were used for decoding, after Gaussian smoothing ($\sigma = 25\ ms$). Training and generalization were performed using data from a single time point with a multiclass (24 classes) linear discriminant analysis classifier, implemented in Matlab (the *fitcdiscr* function), according to the following equation:

$$\hat{y} = \underset{y\ =\ 1\ \ldots\ 24}{argmin} \sum_{k=1}^{24} \hat{P}(k|x)C(y|k)$$

where $\hat{y}$ is the predicted class, k is the sampled class, y is the real class, $\hat{P}(k|x)$ is the posterior probability of class k for observation x and $C(y|k)$ is the cost of classifying an observation as y when the true class is k (i.e. 1 if correct, 0 if wrong). No regularization was used. In addition,

we repeated the procedure for every possible pair of time points between training and generalization - a procedure known as temporal generalization[18] (Figure S2).

V1 responses elicited by non-occluded stimuli had a higher SNR than those elicited by the occluded condition (measured as the ratio between the mean and standard deviation of the response). It has been shown that training on a dataset with higher SNR leads to poor generalization when testing on a dataset with lower SNR[44]. In order to address this issue, when generalizing from the non-occluded to the occluded condition, we conducted extra trial averaging (10 random trials, constant across cross-validation folds) in the test-set of occluded data to match the SNR of V1 responses elicited by the non-occluded stimuli (SNR after stimulus onset was 3.1 for the non-occluded condition, 1.8 for the occluded condition and 3.2 for the matched occluded condition).

To avoid overfitting and to obtain a robust estimate of the generalization performance, we repeated a cross-validation procedure five times (i.e. we trained the decoders on a randomly selected 80% of the trials and tested them on the held-back 20% of trials and repeated this procedure 5 times) and averaged across the accuracies.

Correlation analysis

We used RSA to gauge the similarity between monkey V1 spiking activity and human V1 functional MRI responses, using data from 18 human subjects (see ref.[8] for details on MRI acquisition and preprocessing). For both monkey and human V1, we computed a RDM based on pairwise distances, computed as linear discriminant contrast (LDC, i.e. the cross-validated Mahalanobis distance between either voxel activations (for humans) or recording sites (for monkeys) activations to different stimuli)[45], and we used the same cross-validation scheme discussed above. For human V1, only voxels representing the occluded image region were included, as reported in ref. [8]. To calculate the similarity between each time point in the data of monkey V1 and the average dissimilarity matrix across human subjects, we used Spearman's rank correlation.

Quantification and statistical analysis

Neural decoding

We estimated the statistical significance of decoding with a permutation test. First, we built a permuted null-distribution by repeating the same procedure 100 times and shuffling the labels of the stimuli (keeping the time points for training and generalization the same). Second, we selected the upper tail (highest 30%) of the null distribution, and we fitted a Pareto function to the tail to estimate extreme values in full[46]. Third, we identified the percentile corresponding to each accuracy value in the null-distribution. Finally, where appropriate, we corrected for multiple comparisons using the false discovery rate (FDR) correction[47].

Correlation analysis

To compute the statistical significance of each correlation value, we employed a permutation test: for each time point in the monkey V1 data, we built a null-distribution repeating the same procedure while randomly shuffling the order of cells in the monkey V1 dissimilarity matrix a thousand times. Then, we computed the p-value of each real correlation value by identifying the percentile corresponding to its value in the null-distribution[48]. Finally, we corrected for multiple comparisons (across time-points) using an FDR correction[47,48]. In

addition, we computed the interval of the highest correlation that was theoretically possible, given the noise in the data (i.e., noise ceiling), using the procedure described in ref.[49].

References

1. Cadena, S.A., Denfield, G.H., Walker, E.Y., Gatys, L.A., Tolias, A.S., Bethge, M., and Ecker, A.S. (2019). Deep convolutional models improve predictions of macaque V1 responses to natural images. PLOS Comput Biol *15*, e1006897. 10.1371/journal.pcbi.1006897.
2. Lamme, V.A., and Roelfsema, P.R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. Trends Neurosci *23*, 571–579.
3. Roelfsema, P.R. (2006). Cortical algorithms for perceptual grouping. Annu Rev Neurosci *29*, 203–227. 10.1146/annurev.neuro.29.051605.112939.
4. Muckli, L., and Petro, L.S. (2013). Network interactions: Non-geniculate input to V1. Curr Opin Neurobiol *23*, 195–201. 10.1016/j.conb.2013.01.020.
5. Muckli, L., de Martino, F., Vizioli, L., Petro, L.S., Smith, F.W., Ugurbil, K., Goebel, R., and Yacoub, E. (2015). Contextual Feedback to Superficial Layers of V1. Curr Biol *25*, 2690–2695. 10.1016/j.cub.2015.08.057.
6. Sugita, Y. (1999). Grouping of image fragments in primary visual cortex. Nature *401*, 269–272. 10.1038/45785.
7. Keller, A.J., Roth, M.M., and Scanziani, M. (2020). Feedback generates a second receptive field in neurons of the visual cortex. Nature *582*, 545–549. 10.1038/s41586-020-2319-4.
8. Morgan, A.T., Petro, L.S., and Muckli, L. (2019). Scene Representations Conveyed by Cortical Feedback to Early Visual Cortex Can Be Described by Line Drawings. Journal of Neuroscience. 10.1523/JNEUROSCI.0852-19.2019.
9. Revina, Y., Petro, L.S., and Muckli, L. (2018). Cortical feedback signals generalise across different spatial frequencies of feedforward inputs. Neuroimage *180*, 280–290. 10.1016/j.neuroimage.2017.09.047.
10. Smith, F.W., and Muckli, L. (2010). Nonstimulated early visual areas carry information about surrounding context. Proc Natl Acad Sci U S A *107*, 20099–20103. 10.1073/pnas.1000233107.
11. Self, M.W., van Kerkoerle, T., Goebel, R., and Roelfsema, P.R. (2019). Benchmarking laminar fMRI: Neuronal spiking and synaptic activity during top-down and bottom-up processing in the different layers of cortex. Neuroimage *197*, 806–817. 10.1016/j.neuroimage.2017.06.045.
12. Roelfsema, P.R., and Houtkamp, R. (2011). Incremental grouping of image elements in vision. Atten Percept Psychophys *73*, 2542–2572. 10.3758/s13414-011-0200-0.
13. Roelfsema, P.R., and de Lange, F.P. (2016). Early Visual Cortex as a Multiscale Cognitive Blackboard. Annu Rev Vis Sci *2*, 131–151. 10.1146/annurev-vision-111815-114443.
14. Zhou, H., Friedman, H.S., and von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. J Neurosci *20*, 6594–6611.
15. Williford, J.R., and von der Heydt, R. (2016). Figure-Ground Organization in Visual Cortex for Natural Scenes. eNeuro *3*. 10.1523/ENEURO.0127-16.2016.
16. Poort, J., Raudies, F., Wannig, A., Lamme, V.A.F., 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. 10.1016/j.neuron.2012.04.032.

17. Angelucci, A., and Bullier, J. (2003). Reaching beyond the classical receptive field of V1 neurons: Horizontal or feedback axons? J Physiol Paris *97*, 141–154. 10.1016/j.jphysparis.2003.09.001.
18. King, J.R., and Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. Trends Cogn Sci *18*, 203–210. 10.1016/j.tics.2014.01.002.
19. Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis–connecting the branches of systems neuroscience. Front Syst Neurosci *2*.
20. Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., and Bandettini, P.A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron *60*, 1126–1141.
21. 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. 10.1093/cercor/bhw235.
22. Papale, P., Betta, M., Handjaras, G., Malfatti, G., Cecchetti, L., Rampinini, A., Pietrini, P., Ricciardi, E., Turella, L., and Leo, A. (2019). Common spatiotemporal processing of visual features shapes object representation. Sci Rep *9*, 1–8. 10.1038/s41598-019-43956-3.
23. Hubel, D.H., and Wiesel, T.N. (1974). Uniformity of monkey striate cortex: A parallel relationship between field size, scatter, and magnification factor. Journal of Comparative Neurology *158*, 295–305. 10.1002/cne.901580305.
24. Rockland, K.S., and Lund, J.S. (1983). Intrinsic laminar lattice connections in primate visual cortex. J Comp Neurol *216*, 303–318. 10.1002/cne.902160307.
25. Angelucci, A., and Bressloff, P.C. (2006). Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. Prog Brain Res *154*, 93–120. 10.1016/S0079-6123(06)54005-1.
26. Hirsch, J.A., and Gilbert, C.D. (1991). Synaptic physiology of horizontal connections in the cat's visual cortex. Journal of Neuroscience *11*, 1800–1809. 10.1523/jneurosci.11-06-01800.1991.
27. Nauhaus, I., Busse, L., Carandini, M., and Ringach, D.L. (2009). Stimulus contrast modulates functional connectivity in visual cortex. Nat Neurosci *12*, 70–76. 10.1038/nn.2232.
28. Benucci, A., Frazor, R.A., and Carandini, M. (2007). Standing Waves and Traveling Waves Distinguish Two Circuits in Visual Cortex. Neuron *55*, 103–117. 10.1016/j.neuron.2007.06.017.
29. Nelson, D.A., and Katz, L.C. (1995). Emergence of functional circuits in ferret visual cortex visualized by optical imaging. Neuron *15*, 23–34. 10.1016/0896-6273(95)90061-6.
30. Pooresmaeili, A., and Roelfsema, P.R. (2014). A growth-cone model for the spread of object-based attention during contour grouping. Curr Biol *24*, 2869–2877. 10.1016/j.cub.2014.10.007.
31. Girard, P., Hupé, J.M., and Bullier, J. (2001). Feedforward and Feedback Connections Between Areas V1 and V2 of the Monkey Have Similar Rapid Conduction Velocities. J Neurophysiol *85*, 1328–1331. 10.1152/jn.2001.85.3.1328.

32. Kirchberger, L., Mukherjee, S., Self, M.W., and Roelfsema, P.R. (2023). Contextual drive of neuronal responses in mouse V1 in the absence of feedforward input. Sci Adv *9*. 10.1126/sciadv.add2498.
33. Walker, E.Y., Sinz, F.H., Cobos, E., Muhammad, T., Froudarakis, E., Fahey, P.G., Ecker, A.S., Reimer, J., Pitkow, X., and Tolias, A.S. (2019). Inception loops discover what excites neurons most using deep predictive models. Nat Neurosci. 10.1038/s41593-019-0517-x.
34. Bashivan, P., Kar, K., and DiCarlo, J.J. (2019). Neural population control via deep image synthesis. Science (1979). 10.1126/science.aav9436.
35. Papale, P., Zuiderbaan, W., Teeuwen, R.R.M., Gilhuis, A., Self, M.W., Roelfsema, P.R., and Dumoulin, S.O. (2021). The influence of objecthood on the representation of natural images in the visual cortex. bioRxiv, 2021.09.21.461209. 10.1101/2021.09.21.461209.
36. Self, M.W., Peters, J.C., Possel, J.K., Reithler, J., Goebel, R., Ris, P., Jeurissen, D., Reddy, L., Claus, S., Baayen, J.C., et al. (2016). The Effects of Context and Attention on Spiking Activity in Human Early Visual Cortex. PLoS Biol *14*, e1002420. 10.1371/journal.pbio.1002420.
37. Ricciardi, E., Papale, P., Cecchetti, L., and Pietrini, P. (2020). Does (lack of) sight matter for V1? New light from the study of the blind brain. Neurosci Biobehav Rev *118*, 1–2. 10.1016/j.neubiorev.2020.07.014.
38. Vetter, P., Bola, Ł., Reich, L., Bennett, M., Muckli, L., and Amedi, A. (2020). Decoding Natural Sounds in Early "Visual" Cortex of Congenitally Blind Individuals. Current Biology *30*, 3039-3044.e2. 10.1016/j.cub.2020.05.071.
39. Chen, X., Possel, J.K., Wacongne, C., van Ham, A.F., Klink, P.C., and Roelfsema, P.R. (2017). 3D printing and modelling of customized implants and surgical guides for non-human primates. J Neurosci Methods *286*, 38–55. 10.1016/j.jneumeth.2017.05.013.
40. Super, H., and Roelfsema, P.R. (2005). Chronic multiunit recordings in behaving animals: advantages and limitations. Prog Brain Res *147*, 263–282. 10.1016/S0079-6123(04)47020-4.
41. Trautmann, E.M., Stavisky, S.D., Lahiri, S., Ames, K.C., Kaufman, M.T., O'Shea, D.J., Vyas, S., Sun, X., Ryu, S.I., Ganguli, S., et al. (2019). Accurate Estimation of Neural Population Dynamics without Spike Sorting. Neuron *103*, 292-308 e4. 10.1016/j.neuron.2019.05.003.
42. Togt, C. van der, Klink, C., Papale, P., and Teeuwen, R. (2022). VisionandCognition/Tracker: Public Release. 10.5281/ZENODO.6489014.
43. Roelfsema, P.R., Khayat, P.S., and Spekreijse, H. (2003). Subtask sequencing in the primary visual cortex. Proc Natl Acad Sci U S A *100*, 5467–5472. 10.1073/pnas.0431051100.
44. van den Hurk, J., and op de Beeck, H. (2019). Generalization asymmetry in multivariate cross-classification: When representation A generalizes better to representation B than B to A. bioRxiv, 592410. 10.1101/592410.
45. Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., and Diedrichsen, J. (2015). Reliability of dissimilarity measures for multi-voxel pattern analysis. Neuroimage.
46. Knijnenburg, T.A., Wessels, L.F.A., Reinders, M.J.T., and Shmulevich, I. (2009). Fewer permutations, more accurate P-values. Bioinformatics *25*, i161–i168. 10.1093/bioinformatics/btp211.

47. Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency.
48. Papale, P., Leo, A., Cecchetti, L., Handjaras, G., Kay, K.N., Pietrini, P., and Ricciardi, E. (2018). Foreground-background segmentation revealed during natural image viewing. eNeuro *5*. 10.1523/ENEURO.0075-18.2018.
49. Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., and Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. PLoS Comput. Biol *10*, e1003553.