

University of Groningen

Emerging perception

Nordhjem, Barbara

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2017

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Nordhjem, B. (2017). *Emerging perception: Tracking the process of visual object recognition*. Rijksuniversiteit Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

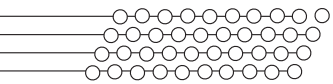
Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

5

Changes in V1
response maps
following object
recognition:
methods and
preliminary
analysis



Abstract

Comprehending the world necessitates the reliable and efficient segmenting of complex visual scenes into coherent objects that can be acted upon. This requires the integration of sensory data and information on behavioral relevance. Where in the human brain this integration into priority occurs remains to be determined. In the present fMRI study, we tested the hypothesis that V1 represents priority by evaluating its response during an object recognition task. Participants viewed emerging images (EIs), similar to the famous Gestalt image of a Dalmatian, for which the recognition process is extended. For such stimuli, the percept changes dramatically within a trial, even though the physical stimulus remains the same. Therefore, our paradigm provides insight into the neuronal responses associated with recognition. For the analysis, we developed a methodological approach combining population receptive field (pRF) mapping with blood oxygen level-dependent (BOLD) responses to chart spatially detailed maps of cortical activity. For many active pRFs, the responses were modulated following recognition despite the physical stimulus remaining the same. A possible explanation for such changes is that V1 activity represents priority and not just the visual input.

Introduction

The human visual system has the remarkable ability to group individual elements together and organize them into coherent and meaningful objects. Selecting relevant information and organizing it into objects, relies on the sensory input as well as experience acquired through daily encounters with the visual world. Hence, stable and versatile object recognition requires both “bottom-up” sensory information and “top-down” influences such as prior knowledge, expectations, and goals. The combined contribution of visual input and top-down influences can be described as priority (Bisley & Goldberg, 2010; Fecteau & Munoz, 2006; Itti & Koch, 2001; Serences & Yantis, 2006). Priority maps in the cortex are thought to represent the degree of attention directed towards spatial locations based on sensory data and behavioral relevance. Hence, specific parts of the visual field will receive higher priority and focused attention. However, where and how top-down and bottom-up information is integrated during object recognition is still unknown.

The primary visual cortex (V1) is a likely candidate for integration of top-down and bottom-up information (Gilbert & Li, 2013; Hochstein & Ahissar, 2002; Lee, Mumford, Romero, & Lamme, 1998; Petro, Vizioli, & Muckli, 2014). Anatomically, V1 receives feedforward projections that carry sensory information as well as feedback projections from higher-order cortical areas (Douglas et al., 2007). V1 also contains the neural mechanisms to form a detailed retinotopic map of the visual field that higher-order areas can act upon to select a subset of the incoming visual information. As not all sections of complex scenes or images are similarly critical for recognition, this model predicts that recognition will result in both increments and decrements of the V1 signals during object recognition.

The results of previous fMRI studies into recognition have been diverse, with evidence for both decreased V1 activity, e.g. due to suppression (Fang, Kersten, & Murray, 2008; Murray et al., 2002), and increased V1 activity, e.g. due to grouping (Altmann, Bu-lthoff, & Kourtzi, 2003; Meng, Remus, & Tong, 2005). A particular reason for these contradicting results may lie in these studies averaging across entire regions of interest (ROIs), thereby neglecting that recognition may simultaneously require the enhancement of responses to relevant information and suppression of responses to irrelevant information. Averaging across regions severely hampers one’s ability to uncover evidence for simultaneously ongoing yet opposing processes within a region.

Indeed, studies using illusory contours have found evidence for both inhibitory and excitatory activations taking place within V1 (Kok, Bains, van Mourik, Norris, & de Lange, 2016; Kok & De Lange, 2014). While such contours are relatively simple stimuli, these studies suggest that V1 may indeed play a substantial role in more complex object recognition, which is often considered a process taking place in higher-order regions (i.e. the inferotemporal cortex).

Typically, previous studies have used alterations of the physical input to induce the transition from unrecognized to recognized object. However, this manipulation induces undesirable correlations between the physical alteration and the perceptual transition. Moreover, previous studies have used simple shapes or geometrical figures. To overcome these limitations, in the present study we used emerging images (EIs): computer-synthesized stimuli with emergent features (Mitra, Chu, Lee, & Wolf, 2009) similar to the famous Gestalt image of a Dalmatian in a sun-spotted garden (photographed by R. C. James). The major advantage of using these images is that the perceptual experience transitions dramatically in the absence of a physical change to the visual stimulation: what initially appears to be a

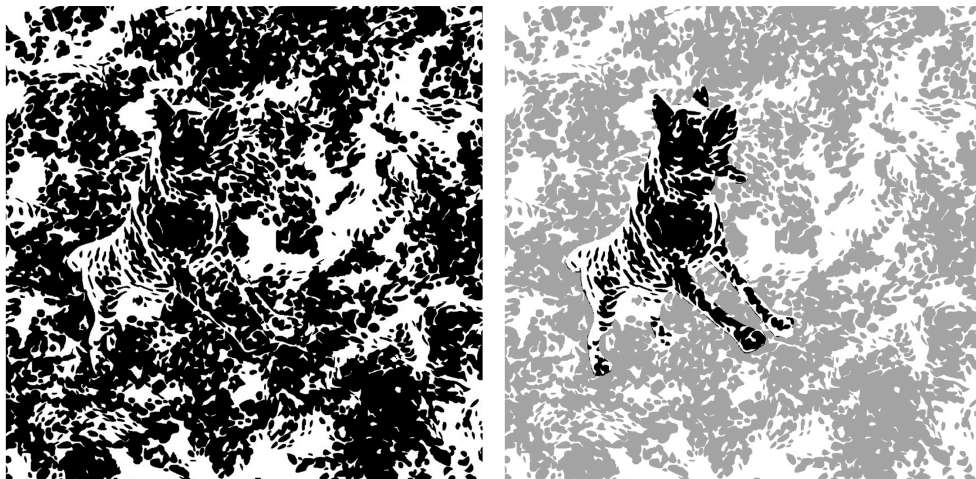
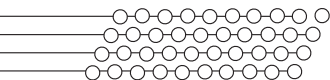


Figure 5.1: Example of an emerging image (EI). The original EI with a hidden dog is shown to the left, and the same image with the dog against a lighter background is shown on the right.

meaningless collection of black and white patches eventually turns out to contain a recognizable object (Nordhjem et al., 2015). Consequently, the sensory input to the early visual cortex remains identical, while the observer's percept changes. Therefore, studying V1 responses using EIs provides a unique window on the role of this region in object recognition. An additional advantage of using EIs is that the recognition process is extended in time (Nordhjem et al., 2015), making this usually very fast process accessible for study by the relatively temporally slow fMRI technique.

To study the potential involvement of V1 in object recognition, we compared spatially detailed maps of blood oxygen level-dependent (BOLD) activity in V1 before and after the recognition of EIs. To do so, we first used fMRI to create maps of the neuronal population receptive fields (pRFs) present in V1 (Dumoulin & Wandell, 2008). Subsequently, these pRF maps were used to project the pre- and post-recognition BOLD activity in V1 onto the visual field, providing detailed image-based maps of recognition-related BOLD modulations in the absence of any change to the physical visual input. We hypothesized the simultaneous presence of both increments and decrements in V1 activity associated with object recognition. Specifically, we expected increased activity in pRF locations corresponding to the object, and decreased activity in pRF locations surrounding it.

5.2 Methods

5.2.1 Participants

Eight healthy right-handed participants with an average age of 24.5 years participated (six women). All participants reported normal or corrected-to-normal vision. The participants were recruited via an advertisement and rewarded with 10 euros per hour for participation. All participants understood the instructions and were able to recognize an example EI prior to scanning.

5.2.2 Ethics statement

The Medical Ethical Committee of the University Medical Center Groningen approved the present study. All participants signed a consent form prior to the study. Participants were informed that the experiment was voluntary and that they could terminate their participation at any time.

5.2.3 Data acquisition

Participants were scanned using a Philips 3 Tesla MRI scanner (Philips, Best, the Netherlands) at the Neuroimaging Centre in Groningen, the Netherlands. High-resolution T1-weighted structural images were acquired at 3 Tesla using a six-channel head coil at a resolution of 1 x1x1 mm³ isotropic voxels, with a field of view of 256x256x170 mm. The TR was 9.00 ms, TE was 3.54 ms, axial orientation. The volume orientation differed between participants, though in all cases it was approximately parallel to the calcarine sulcus.

The retinotopy scan had a repetition time (TR) of 1.5 s and an echo time (TE) of 30 ms, and 24 slices were collected with a resolution of 2.5x2.5x2.5 mm³ isotropic voxels. The retinotopy scan lasted 210 s per run, and 132 volumes were collected. Each run was repeated six times. The EI session had a TR of 2 s, a TE of 30 ms, and 36 slices were acquired. In total, 300 volumes were collected.

Stimuli were presented on a 24 inch BOLDscreen, an fMRI-compatible LCD screen with the dimensions of 620 x 445 mm and a refresh rate of 60 Hz. The resolution was set to 1920x1200 pixels. The distance between the subject and the screen was approximately 120 cm. An Eyelink 1000 eye tracker was used in the EI session to ensure stable fixations and track eye movements during the periods of free viewing (not reported on in this chapter). Participants responded using an MRI-compatible button box. Behavioral responses and eye movement data were collected using MATLAB (MathWorks, Natick, MA, USA) with the Psychophysics toolbox (PTB-3; Brainard, 1997; Pelli, 1997) in combination with the Eyelink Toolbox (Cornelissen, Peters, & Palmer, 2002).

5.2.4 Procedure

All participants came for two fMRI sessions on separate days. During the first session, an anatomical scan and retinotopic mapping were carried out, while during the second session the EI experiment was performed. Each session took approximately one and a half hours including preparation time, debriefing, and actual scan time.

5.2.5 Retinotopic mapping

For the retinotopy, participants were instructed to focus on the colored dot in the middle of the screen and not to track the bar moving across their visual field. Participants were asked to press a button as soon as the fixation dot changed color to ensure stable fixation and attention to the stimulus. During the first fMRI session, participants were shown a moving bar stimulus to map the visual regions using the pRF method (Dumoulin & Wandell, 2008).

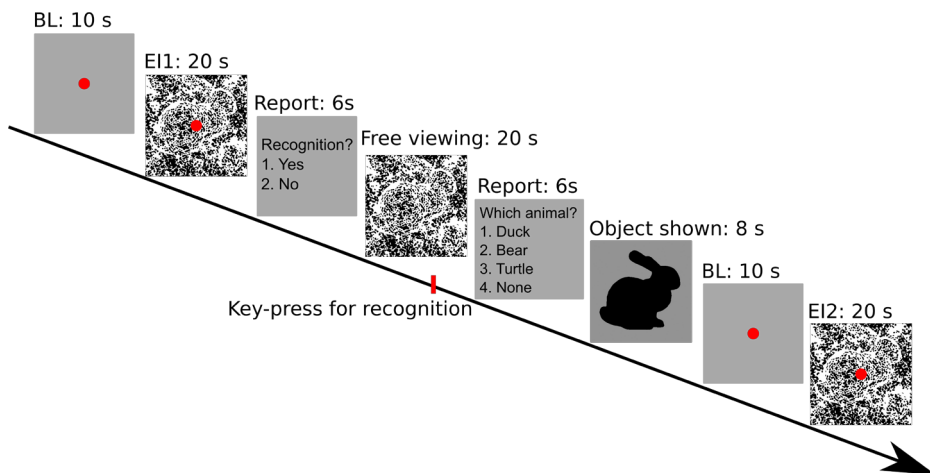
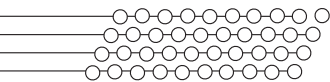


Figure 5.2: Schematic representation of the paradigm. Each trial consisted of the baseline (BL) and the first EI with a central fixation point (EI1), followed by a report during which the participant could indicate by key-press if the EI was already recognized. Subsequently, the hidden object was shown, baseline was collected again, and finally the same EI was presented again (EI2).

5.2.6 EI experiment

For the EI session, the participants were instructed prior to scanning and shown a sample trial on a laptop. The instructions within the overall experiment were to look at the images and follow the instructions on the screen. All EIs were presented twice (EI1 and EI2) and participants were requested to maintain fixation on a circle shown at the center of the screen (Figure 5.2). Immediately after EI1 had been shown, the participants were asked to indicate whether they had been able to recognize an object. Recognition was defined as the ability to perceive a shape that could be categorized as a specific object, and a few examples were given before scanning started (e.g. perceiving something or an animal should not be indicated as recognition, while perceiving a dog or a wolf would count as recognition). Subsequently, participants were allowed to explore the image by freely viewing it; the combined eye tracking and fMRI results are not presented here. During the free viewing, participants were asked to make a key-press if they recognized an object. After the free viewing, participants were then given four choices to indicate what they had recognized (this was always three different animals and the option that nothing was recognized). Finally, an image clearly showing the embedded shape was shown and EI2 was presented.

5.2.7 EI stimuli

The EIs were all created and validated in a previous study (see Nordhjem et al., 2015 for details about image creation). Fifteen images with an emerging object and five similarly textured images (nonsense1-5) without a hidden object were included. The images were viewed at a diameter of 10.2° to match the image size with the stimuli used for retinotopic mapping. The EIs were generated using an algorithm that mapped black splats onto a 3D object that itself was embedded in a background with similar appearance but lacking the structure of the location of the object (Mitra et al., 2009). The EI stimuli were presented with a central red fixation point.

5.3 Data analysis

5.3.1 Preprocessing

Boundaries between gray matter and white matter were detected using Freesurfer (<http://surfer.nmr.mgh.harvard.edu/>) based on the anatomical T1-weighted image. After the automatic segmentation, each slice was manually checked and corrected using ITK-snap (<http://www.itksnap.org/>). Further preprocessing was carried out using mrVista (<http://vistalab.stanford.edu/software/>). This included motion correction of the functional scans (within and between scan correction) (Nestares & Heeger, 2000) and alignment of the functional data to the anatomical T1-weighted image. The cortical surface was reconstructed and rendered as a 3D surface for visualization, which was used to draw the ROIs (Wandell, Chial, & Backus, 2000).

5.3.2 Analysis overview

Following the preprocessing, the analysis was conducted in several steps (Figure 5.3). The model-based pRF method was used to estimate visual field maps with the data from the retinotopy session (Dumoulin & Wandell, 2008). A general linear model (GLM) was used to contrast activity before and after recognition; this was done at the level of each subject for each stimulus with the data from the EI experiment. The pRF parameters obtained with retinotopic mapping were projected onto the visual field (image space) to obtain visual field coverage maps. Then, the BOLD responses obtained from a GLM analysis were projected onto the visual field using the pRF parameters. Thus, priority maps were created for each individual observer and each EI.

5.3.3 pRF analysis and definition of ROIs

The pRF of a voxel is defined as the location in visual space that maximally stimulates it (Dumoulin & Wandell 2008). Briefly described, the pRF analysis generates predictions by varying location (x, y) and spread (σ) of a Gaussian for each voxel until best fit is reached (minimum residual sum of squares). The fMRI time series for each voxel is computed by convolving the pRF model with the stimulus matrix and the BOLD hemodynamic response function (HRF) (Boynton, Engel, Glover, & Heeger, 1996; Friston et al., 1998; Glover, 1999; Worsley et al., 2002). This analysis is described in detail in Dumoulin and Wandell (2008). The pRF models were used to estimate the pRF parameters for each voxel and create eccentricity and polar angle maps to delineate V1. We defined the ROIs individually for each subject. Polar- angle and eccentricity maps were created by projecting the pRF estimates onto the 3D model of the gray-white matter border for each subject. V1 was obtained by outlining the borders on the 3D model for each hemisphere and then combined (Figure 5.3).

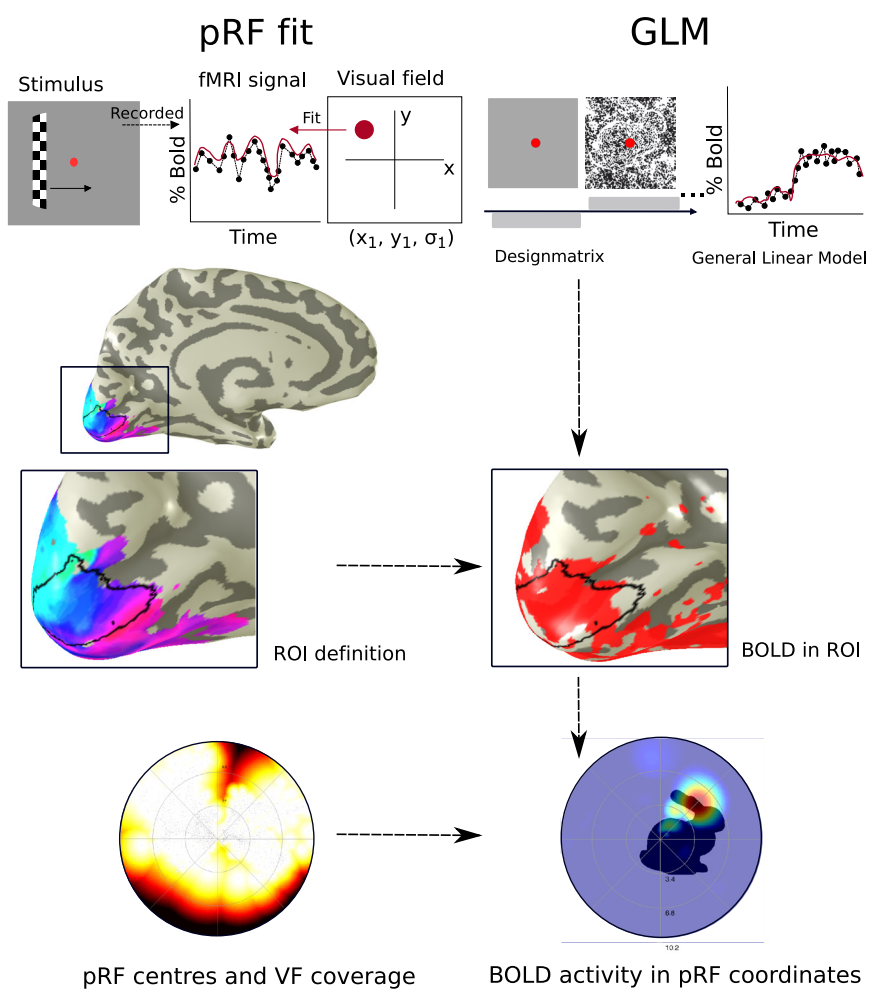
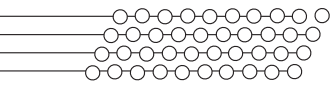


Figure 5.3: Overview of the analysis steps. The pRF analysis (left panel) was conducted to define V1 as an ROI for each participant and to obtain the pRF location and visual field coverage for each voxel. A GLM analysis (right panel) was performed to correlate BOLD responses to the stimuli per voxel. Finally, BOLD activity within the ROI was projected on the visual stimulus using pRF coordinates and size (i.e. VF coverage).

5.3.4 GLM analysis of the EI experiment

Activity for each stimulus before (EI1) and after recognition (EI2) was analyzed using a GLM in mrVista. For each stimulus, a parameter file was created to specify the design matrix. For each subject s and each EI object i , the onsets of EI1 $_{s,i}$, EI2 $_{s,i}$ and Baseline (BL) were modeled, while instructions between stimulus presentation and the rest of the EIs were modeled as regressors-of-no interest. We included the following contrasts for each subject s per image i : EI1 $_{s,i} > BL$, EI2 $_{s,i} > BL$, and EI2 $_{s,i} > EI1_{s,i}$. Contrast effect sizes (ces ; arbitrary units) were computed and stored in matrices for further analysis.

5.3.5 BOLD coverage maps

The visual field coverage was estimated for V1 in all subjects. Each coverage plot visualized the locations in the visual field that evoked significant activations in the voxels of a specific visual field map. The visual field was estimated based on the full pRF model. The locations of existing pRF centers were mapped as a binary image, and the coverage was estimated by combining the pRF size and visual field coverage parameters. Hence, for each voxel in the ROI, its estimated 2D pRF model was projected onto the visual field (stimulus referred). If a location was covered by several pRFs, the one with the highest weight was plotted (maximum profile). For each individual observer, a V1 visual field coverage plot was created that combined pRFs estimated for the left and right hemispheres.

We then charted the BOLD responses in visual field coordinates for EI1, EI2, and the difference between them: ΔEI . This analysis was conducted to show the relative importance of locations in the visual field (and thus image) in causing recognition-related modulations. For each individual participant and stimulus, we created maps of the BOLD activity in visual field coordinates:

$$Max_n(ces_i \cdot g(x0_i, y0_i, \sigma_i))$$

Equation 5.1

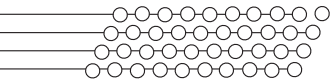
Where n denotes the number of voxels in the given ROI, ces_i is the contrast effect size, and $g(x0_i, y0_i, \sigma_i)$ is the 2D Gaussian with peak coordinates $(x0, y0)$ and standard deviation (σ) . Only voxels from the pRF model with a variance explained (VE) threshold of 0.5 were included in the analysis. Because single voxels with a very large RF size could distort the results, we used a median filter that smoothed the pRF size parameter of each voxel with its two nearest neighbors and replaced that voxel's size parameter with the median of the three voxel sizes.

5.3.6 Individual offsets

We computed individual offsets to account for the possibility that some voxels might just always be more active, e.g. due to a foveal bias. The offset for each subject was calculated by averaging the contrast effect sizes for all EIs ($s = 1 \dots N$ subject, i image).

$$\Theta = \frac{\sum EI1_s, EI2_s}{N_{EI1, EI2}}$$

Equation 5.2



Following this, the maps were corrected for the offsets for EI1 and EI2. Notice that the subject offset (Θ) is subtracted from each image. The multiplication by the offset was included to avoid spurious results from non-prioritized regions (i.e. very large differences in areas where there was no effect of the current image).

$$\sum_{s,i} \Theta_s (EI1_{s,i} - \Theta_s)$$

Equation 5.3

$$\sum_{s,i} \Theta_s (EI2_{s,i} - \Theta_s)$$

Equation 5.4

The effect of recognition, defined as the relative difference between EI2 and EI1, was corrected for the individual offsets as follows:

$$\sum_{s,i} \Theta_s \Delta EI_{s,i}$$

Equation 5.5

5.3.7 Group analysis

To determine the effect of recognition, we plotted the mean BOLD activation for each EI across observers $EI2_i > EI1_i$. All data that entered the group analysis were corrected for individual offsets as described in the previous section. Permutation tests were carried out to assess which locations in the VF (and hence on each EI) were modulated significantly due to recognition. The null hypothesis was that there is no effect of recognition and therefore no difference between E1 and E2. The labels E1 and E2 were permuted, which is equivalent to permuting their sign. For each image, 256 permutations were carried out (which is the maximum number of permutations possible given the number of participants). The group results are shown per image in VF coordinates with contour lines showing the areas of the EI where activity deviated significantly from the null hypothesis. The statistical threshold was set at the 95th percentile of the permuted distribution.

5.3.8 Visualization of saliency

To investigate whether brain activity found for EI1 and EI2 could be explained by low-level image properties, we visualized saliency maps. We used two computational models of saliency: the classic saliency model (Itti, Koch, & Niebur, 1998) and the Graph-Based Visual Saliency (GBVS) model (Harel, Koch, & Perona, 2006). We used the GBVS Matlab toolbox by J. Harel, which includes both saliency models (<http://www.vision.caltech.edu/harel/share/gbvs.php>). Saliency maps are shown in visual field coordinates.

5.4 Results

We used fMRI to assess changes in V1 activity during the recognition of EIs. By using coverage maps, we focus on changes in the spatial distribution of the signals. Results are reported across observers. We refer to the supplementary material for recognition performances and additional plots of cortical activity.

5.4.1 Cortical responses to E1 and E2

We plotted the average cortical response maps across participants for E1 and E2; we will refer to these maps as BOLD coverage maps (BCM) 1 and 2. Our hypothesis predicted that the V1 responses in BCM1 would primarily reflect the bottom-up visual input and be largely driven by contrast (i.e. saliency), while the responses to BCM2 would reflect both the visual input as well as prioritized parts of the stimulus (corresponding to (parts of) the recognized object).

Figure 5.4 shows BCM1 and BCM2 for four stimuli. For both the frog and the horse image, there were areas with very similar activity close to the fovea. In itself, this is not so surprising since the visual input was kept the same and participants held central fixation. Hence, the activity in both maps appears for a large part to have been driven by the visual input. For the gorilla and the wolf images, however, there were larger differences between BCM1 and BCM2. For instance, for the gorilla, in BCM1 there was an area of deactivation at the location of the head which became active in BCM2. For the wolf, the reverse pattern was seen compared to the gorilla, namely a large area with an increased response in BCM1 became suppressed in BCM2.

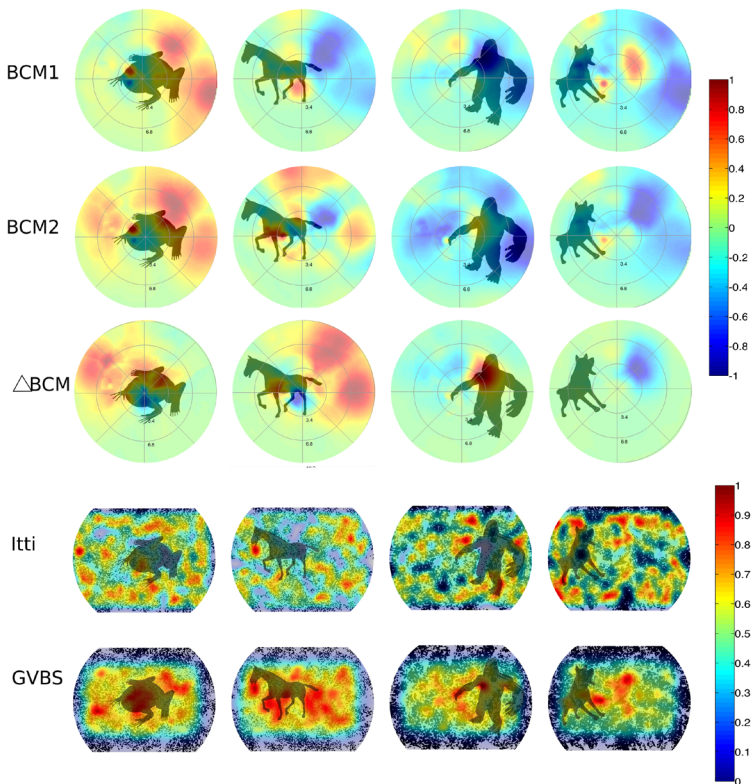
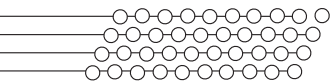


Figure 5.4: BOLD coverage maps and saliency maps. Rows 1-3 show BOLD visual field coverage maps before (BCM1) and after (BCM2) recognition averaged across participants ($n = 8$). The color map represents the normalized group BOLD responses (a. u.). Rows 4-5 show saliency maps computed for the Itti and GBVS models. The object area is shaded for the purpose of illustration. The color map represents normalized saliency scores.



5.4.2 Changes in cortical activity

Our goal was to map changes in cortical activity related to object recognition while the visual input remained the same. To do so, for each observer we computed the difference between BCM1 and BCM2: ΔBCM . The average ΔBCM maps reveal the spatial differences in activity between BCM1 and BCM2. They thus reveal regions of V1 that differentially responded following the recognition of the objects. Figure 5.4 (row 3) shows that for the frog and gorilla image, such regions were located at visual field locations both inside and near the edges of the objects. However, for the horse image, large differences can be seen outside the object. For the wolf image, on the other hand, the ΔBCM shows little specific differential activity. Hence, we do find substantial spatial differences in V1 responses before and after recognition despite the fact that the physical stimuli did not change. Locations of increased BOLD responses corresponded to those of the object as well as to locations outside of the objects (see supplementary material for additional BCMs).

5.4.3 Relating saliency to BOLD responses

To qualitatively examine the role of the physical stimulus in the cortical responses, we created saliency maps for all EIs using two models (referred to as Itti and GBVS, respectively). The objective was to compare whether salient parts of the EIs would overlap with the locations of higher cortical activity in BCM1 and BCM2. The saliency maps are shown in Figure 5.4 (bottom two rows).

In general, neither saliency model revealed the entire object in any of the EIs as a single salient region. The Itti model revealed many small salient areas scattered across the EI, whereas the GBVS model revealed fewer but larger salient areas. For the frog image, there appears to be some overlap between the cortical activity seen for BCM1 and BCM2 and the two salient areas just behind the frog in the GBVS map. However, there is also a salient area in the GBVS map that is not reflected in the activity pattern in the BCMs. For the horse image, there is no obvious relationship between the saliency map and the BCM activity. For the gorilla image, in contrast to the horse, there is correspondence between the salient location near the head as revealed by the GBVS map, and the activity in the ΔBCM . However, for many other locations, no such relationship existed. For the wolf image, the Itti saliency map is scattered and does not correspond particularly well to either of the BCMs. However, the GBVS map shows two salient areas that also show increased activity in BCM1. The same highly salient area shows decreased activity for BCM2, indicating that responses to this area of the image became suppressed. Hence, it appears that there is occasional correspondence between saliency as modeled with GBVS and cortical activity in the BCM plots, especially prior to recognition (BCM1). For the Itti model, however, the relationship seems absent.

5.5 Discussion

We found that V1 responses change following the recognition of an object. In the absence of physical changes to the stimulus, this suggests that the origins of the local increases in signal are associated with the recognition process itself.

Visual object recognition and its underlying processes such as grouping and image segregation are still poorly understood. The classic feedforward modular view suggests that V1 extracts simple features

such as edges, while more complex shape representation takes place in higher-order extrastriate areas (DiCarlo, Zoccolan, & Rust, 2012). Here, we show that V1 activity is modulated by object recognition in the absence of any changes to the visual stimuli. One interpretation of the increased V1 activity is that it represents the priority that certain visual locations have for the recognition process. We will discuss these matters in more detail below.

5.5.1 V1 responses are modulated by object recognition

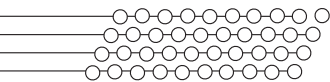
We expected increased activity in pRF locations corresponding to the object, and decreased activity in pRF locations surrounding it. However, the pattern of the results we obtained was less clear. For some of the stimuli, the difference in V1 response before (BCM1) and after recognition (BCM2) revealed that object recognition was associated with increased activation of pRFs at locations corresponding to the object. However, in other cases, the activations and deactivations showed no clear relationship to the object present. Hence, while we did find recognition-related V1 modulations, the degree to which these reflect perceptual changes remains unclear.

Grouping and object recognition may depend on the interplay between the early visual cortex and higher-order regions that respond to object shape. Therefore, retinotopically organized regions with smaller RF sizes such as V1, and non-retinotopic areas with large RF sizes have complimentary roles. One possibility is that bidirectional connectivity between the early visual and higher-order areas is required to perceive both the global shapes and their finer details (Lee et al., 1998). However, our research indicates that object recognition does not modulate feedback effective connectivity to V1 (Nordhjem et al., 2016).

The question that arises is why these changes in responses take place in V1 and not only in higher-order visual areas. One speculation is that higher-order visual areas with large – spatially invariant – receptive fields are involved in object recognition at a more global and conceptual level, while retinotopically organized visual areas such as V1 enhance and suppress incoming information based on the (expected or inferred) behavioral relevance of specific stimulus regions. In this vein, the modulations at the level of the early visual areas may be part of grouping processes that help reduce complexity. As soon as local elements can be assigned as being part of a texture or as belonging to either the fore- or the background, they may not require further detailed processing by higher-order areas. In the case of EIs, such assignment may not always be fully correct, which could be an explanation for the somewhat variable results.

5.5.2 Variation in BOLD coverage maps across stimuli

For the frog, the gorilla, and somewhat less for the horse EIs, we found increases in V1 activity after recognition corresponding to object locations. However, many Δ BCM modulations showed no clear correspondence to the object locations. This lack of consistency in the maps could reflect substantial variation in the underlying neural recognition processes and strategies across participants. Moreover, for the individual EIs, there were variations in the difficulty with which the object could be segregated from the background. Furthermore, as observers were required to fixate on the center of the image, the eccentricity at which the object was located could play a role in this. There is a higher density of smaller-sized pRFs near the fovea, whereas these become larger and somewhat sparser at higher eccentricity. Hence, to some extent, the BCMs also reflect the quality and quantity of the pRF mapping.



For some of the EIs, image regions outside the object but with a relatively high saliency may have attracted initial attention, which could also be reflected in the BCMs. This could for instance be the case for the wolf image. For this stimulus, a strong response to an area with dense black spots was present before recognition, while after recognition activity in this region was suppressed. Hence, the BCMs may also reflect suppression of distracting information and not only enhancement of relevant information. In support of this possibility, the GBVS saliency model showed a high saliency in approximately the same location where we observed cortical activity in BCM1 and a decrease in activity for BCM2. In any case, this observation supports the notion that the Δ BCM modulations reflect recognition-related activity and not only image statistics.

5.5.3 Limitations

It is challenging to create EIs that are not recognized immediately but that will eventually be by most observers. Compared to other stimuli, e.g. ones inducing illusory contours such as Kaniza shapes and oriented line segments (Lamme, Super, & Spekreijse, 1998), EIs are more complex and perceptually stand out less clearly. Because EIs are perceptually challenging, this could eventually lead to weaker or noisier responses following recognition. Moreover, some images may not have been sufficiently recognizable for some participants. Furthermore, all the objects came in different shapes and sizes, which made our stimuli more complex than previously used geometrical shapes. For that reason, the analysis could only be carried out for one EI at a time. Straightforward averaging across stimuli was not possible, which makes the results somewhat more difficult to interpret in a quantitative manner.

Even though this is a general limitation of fMRI, it is important to note that causality cannot be inferred from the BOLD response maps. It is therefore impossible to determine whether increases in activity during recognition are a prerequisite or a consequence of object recognition. Participants were requested to keep central fixation and responded with a key-press if the color changed. Despite the requirement to attend to the central fixation spot, however, changes in covert attention may have affected the results. After recognition participants may have distributed their attention, or been paying more attention to the object location or specific image features. Finally, in future experiments, it could be advantageous to present the objects at a consistent location and at the same eccentricity. This might facilitate drawing conclusions from multiple stimuli and avoid differences related to the density or reliability of the various pRFs.

5.5.4 Future directions

The EIs all consisted of dense and detailed black splats, and it may require foveal vision to segment the figures from the background. The role of eye movements while recognizing EIs have been treated in the previous chapter, where we showed that fixations are primarily made on the boundaries between objects and their background (Nordhjem et al., 2015). This indicates that inspection of the edges precedes visual recognition. Since we required central fixation, the participants did not freely inspect the details of the stimuli. We made this choice to be able to consistently map the responses prior to and after recognition. In the future, however, it may be possible to allow eye movements and correct the pRF model accordingly (Hummer et al., 2016). This would allow for the integration of the tracking of gaze behavior and brain responses leading up to the moment of recognition.

Cortical activity maps reflecting behavioral relevance offer the possibility of detailed analysis of the neural processes underlying visual perception. A more detailed view of the changes of activity within regions could potentially bring neuroscience closer to answering the question of “how” instead of “where” object recognition is achieved. Whereas the present study was highly explorative, further developments of the approach could complement existing tools in computational visual neuroscience (Wandell & Winawer, 2015).

5.6 Conclusion

We have shown that responses within the V1 are modulated when local features are grouped into global shapes during object recognition. We believe that this map characterizes the responses in V1 that reflect interaction between the early visual cortex and higher-order visual regions. Hence, V1 may enhance the neural responses to prioritized features and suppress those that are of less relevance.

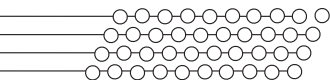
Our results indicate that V1 does not only respond to the visual stimulus, but also includes top-down information from higher-order areas. We have provided a method to project changes in V1 onto the visual field, giving a map of cortical activity in visual field coordinates. Ideally, this method could allow mapping of the locations of attention.

Acknowledgements

This study was done in collaboration with Hinke Halbertsma, Nicolás Gravel, Remco Renken, and Frans W. Cornelissen. We thank Ben Harvey for his insightful suggestions and support with the analysis. We are also grateful for the comments of René Passet and Tim Thomas.

Funding

BN was supported by a grant from the Netherlands Organisation for Scientific Research (NWO Brain & Cognition grant 433-09-233) to FWC. NG was supported by a scholarship from the (Chilean) National Commission for Scientific and Technological Research (BECAS CHILE & Millennium Center for Neuroscience CENEM NC10 001 F).



References

- Bisley, J. W., & Goldberg, M. E. (2010). Attention, Intention, and Priority in the Parietal Lobe. *Annual Review of Neuroscience*, 33(1), 1–21.
- Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *Journal of Neuroscience*, 16(13), 4207–21.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433–436.
- Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The EyeLink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers*, 34(4), 613–617.
- DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, 73(3), 415–34.
- Douglas, R. J., & Martin, K. A. (2007). Mapping the matrix: the ways of neocortex. *Neuron*, 56(2), 226–238.
- Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *NeuroImage*, 39(2), 647–60.
- Fecteau, J. H., & Munoz, D. P. (2006). Saliency, relevance, and firing: a priority map for target selection. *Trends in Cognitive Sciences*, 10(8), 382–390.
- Friston, K. J., Fletcher, P., Josephs, O., Holmes, a, Rugg, M. D., & Turner, R. (1998). Event-related fMRI: characterizing differential responses. *NeuroImage*, 7(1), 30–40.
- Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews. Neuroscience*, 14(5), 350–63.
- Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*, 9(4), 416–29.
- Harel, J., Koch, C., & Perona, P. (2006). Graph-based visual saliency. *Advances in Neural Information Processing Systems*, 19, 545–552.
- Hochstein, S., & Ahissar, M. (2002). View from the Top : Hierarchies and Reverse Hierarchies in the visual system. *Neuron*, 36(5), 791–804.
- Hummer, A., Ritter, M., Tik, M., Ledolter, A. A., Woletz, M., Holder, G. E., ... Windischberger, C. (2016). Eyetracker-based gaze correction for robust mapping of population receptive fields. *NeuroImage*, 142, 211–224.
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews. Neuroscience*, 2(3), 194–203.
- Itti, L., Koch, C., & Niebur, E. (1998). A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11), 1254–1259.
- Kleiner, M., Brainard, D. H., Pelli, D. G., Broussard, C., Wolf, T., & Niehorster, D. (2007). What's new in Psychtoolbox-3? *Perception*, 36, S14.
- Kok, P., Bains, L. J., van Mourik, T., Norris, D. G., & de Lange, F. P. (2016). Selective Activation of the Deep Layers of the Human Primary Visual Cortex by Top-Down Feedback. *Current Biology*, 26, 371–376.
- Kok, P., & De Lange, F. P. (2014). Shape perception simultaneously up- and downregulates neural activity in the primary visual cortex. *Current Biology*, 24(13), 1531–1535.
- Lamme, V., Super, H., & Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. *Current Opinion in Neurobiology*, 8(4), 529–535.
- Lee, T. S., Mumford, D., Romero, R., & Lamme, V. A. (1998). The role of the primary visual cortex in higher level vision. *Vision Research*, 38(15–16), 2429–2454.
- Mitra, N., Chu, H., Lee, T., & Wolf, L. (2009). Emerging images. *ACM Transactions on Graphics*, 28(5), 1–8
- Nestares, O., & Heeger, D. J. (2000). Robust multiresolution alignment of MRI brain volumes. *Magnetic Resonance in Medicine*, 43(5), 705–15.
- Nordhjem, B., Ćurčić-Blake, B., Meppelink, A. M., Renken, R. J., de Jong, B. M., Leenders, K. L., ... Cornelissen, F. W. (2016). Lateral and Medial Ventral Occipitotemporal Regions Interact During the Recognition of Images Revealed from Noise. *Frontiers in Human Neuroscience*, 9, 678.
- Nordhjem, B., Kurman, C. I., Renken, R. J., & Cornelissen, F. W. (2015). Eyes on emergence: Fast detection yet slow recognition of emerging images. *Journal of Vision*, 15(9), 8.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442.
- Petro, L. S., Vizioli, L., & Muckli, L. (2014). Contributions of cortical feedback to sensory processing in primary visual cortex. *Frontiers in Psychology*, 5, 1–8.
- Serences, J. T., & Yantis, S. (2006). Selective visual attention and perceptual coherence. *Trends in Cognitive Sciences*, 10(1), 38–45.
- Wandell, B. A., Chial, S., & Backus, B. T. (2000). Visualization and measurement of the cortical surface. *Journal of Cognitive Neuroscience*, 12(5), 739–52.
- Wandell, B. A., & Winawer, J. (2015). Computational neuroimaging and population receptive fields. *Trends in Cognitive Sciences*, 19(6), 349–357.
- Worsley, K. J., Liao, C. H., Aston, J., Petre, V., Duncan, G. H., Morales, F., & Evans, A. C. (2002). A general statistical analysis for fMRI data. *NeuroImage*, 15(1), 1–15.

Supplementary material

S1. Behavioral results

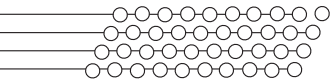
We recorded how many of the EIs were recognized upon the first presentation (EI1) and after the period of free viewing prior to the second presentation (EI2) (Table 5.1). Participants reported whether they had recognized an object (yes or no) at the moment of recognition during free viewing, and were given four choices after the free viewing period.

It proved difficult to find EIs that were sufficiently difficult to not be recognized during the first presentation, but that were recognized correctly after the free viewing condition. A dolphin EI was recognized during EI1 and also during free viewing, and was therefore not a useful stimulus to compare V1 activity before and after recognition. Other objects such as the goat were not recognized correctly by any of the observers and might therefore have been too ambiguous. We chose to focus on the frog, horse, gorilla, and wolf in the further analysis because the majority of the participants did not recognize these stimuli until after the free viewing.

In most cases, recognition was not indicated after EI1 or during free viewing for the images without an object (nonsense 1-5). However, participants often reported seeing an object when given four choices, and answered consistently. For instance, four subjects indicated having seen a whale in one of the nonsense EIs, and six saw a spider in another nonsense EI. Hence, it appears that participants preferred to take a guess instead of indicating that they had seen no object.

Table 5.1: Count of participants who recognized each EI for $n = 7$ (one participant was scanned with a defect button box). EI1 indicates the count of participants who responded that they recognized an object after the first stimulus presentation. FW indicates the count of participants who indicated that they recognized an object during the period of free viewing. Correct response indicates the participants who indicated the correct object when given four choices. The count of participants who responded correctly or saw a similar object is indicated as well.

Image	EI1	FW	Correct response	Correct or similar response
Bear	1	6	2	
Bunny	5	6	1	5 (Cat)
Camel	6	6	2	
Cow	6	6	2	5 (Zebra)
Dolphin	6	6	6	
Elephant	3	6	6	
Flamingo	5	6	0	
Frog	1	6	7	
Goat	7	6	0	
Horse	2	6	3	7 (Gazelle)
Gorilla	1	6	6	
Lion	3	6	0	4 (Donkey)
Panther	6	7	5	
Man	1	6	0	
Wolf	1	5	5	
Nonsense 1	0	0	0	
Nonsense 2	0	0	0	
Nonsense 3	1	0	0	
Nonsense 4	1	1	0	
Nonsense 5	0	1	0	



S2. Individual offsets

Individual offsets are shown for each of the eight subjects in visual field coordinates and normalized to a scale between 0 and 1 in arbitrary units. Strikingly, the offsets are quite heterogeneous: while some subjects showed a clear foveal bias, others on average showed increase in activity at various more peripheral locations.

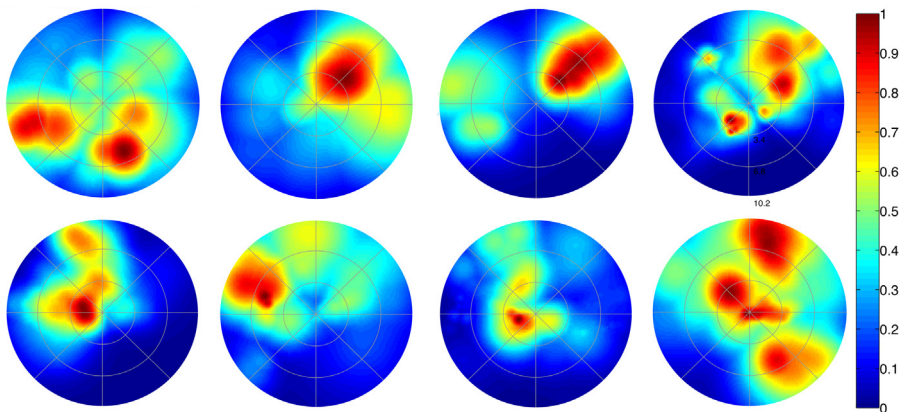


Figure 5.5: Individual offsets

S2. Additional BOLD coverage maps

BOLD coverage maps are shown for the bear, bunny, camel, and nonsense1 image. We did not include these EIs in the main results section because we focused on a subset of images that were not recognized initially, but were successfully recognized after the second presentation.

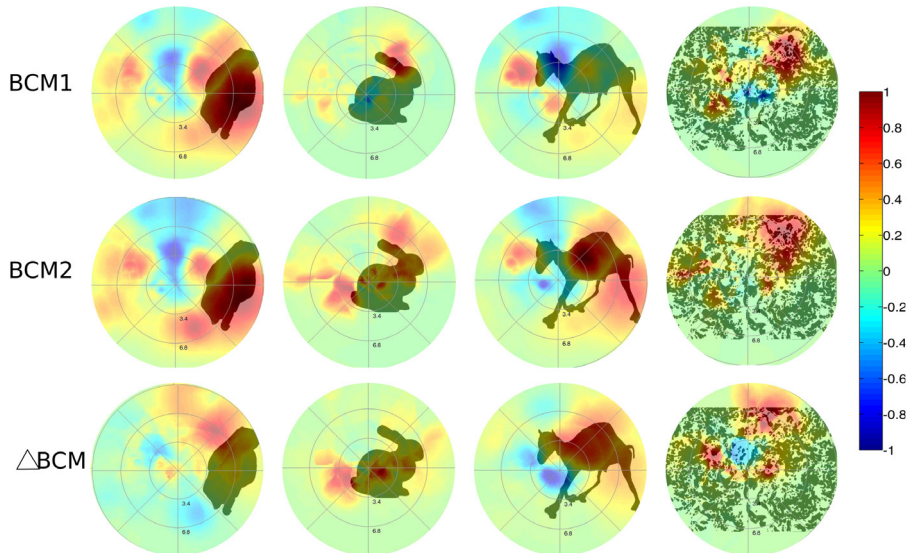


Figure 5.6: Example BOLD coverage maps across participants. The maps show BOLD visual field coverage maps before (BCM1) and after (BCM2) recognition averaged across participants ($n = 8$). The object area is shaded for the purpose of illustration. The color map represents normalized saliency scores.

