Regions of mid-level human visual cortex sensitive to the global coherence of local image patches

Damien J. Mannion[1,2], **Daniel J. Kersten**[1,3], and **Cheryl A. Olman**[1]

[1]Department of Psychology, University of Minnesota, Minneapolis, MN, USA

[2]School of Psychology, University of New South Wales, Sydney, Australia

[3]Department of Brain and Cognitive Engineering, Korea University, Seoul 136-713, Korea

Abstract

The global structural arrangement and spatial layout of the visual environment must be derived from the integration of local signals represented in the lower tiers of the visual system. This interaction between the spatially local and global properties of visual stimulation underlies many of our visual capacities, and how this is achieved in the brain is a central question for visual and cognitive neuroscience. Here, we examine the sensitivity of regions of the posterior human brain to the global coordination of spatially displaced naturalistic image patches. We presented observers with image patches in two circular apertures to the left and right of central fixation, with the patches drawn from either the same (coherent condition) or different (non-coherent condition) extended image. Using functional magnetic resonance imaging (fMRI) at 7T (*n* = 5), we find that global coherence affected signal amplitude in regions of dorsal mid-level cortex. Furthermore, we find that extensive regions of mid-level visual cortex contained information in their local activity pattern that could discriminate coherent and non-coherent stimuli. These findings indicate that the global coordination of local naturalistic image information has important consequences for the processing in human mid-level visual cortex.

Introduction

Visual field selectivity is perhaps the most pronounced response characteristic of neurons in lower tiers of the visual system; a neuron that modulates its activity with great vigour to stimulation within a portion of the visual field will fall silent when the stimulation is moved a short distance away. This receptive field selectivity (Hartline, 1938) distributes the representation of the spatial structure of visual stimulation across a vast neural population, with each neuron influenced from only a restricted *local* part of the visual field. This information must be spatially integrated at higher levels of the visual hierarchy to allow for the recovery of more *global* aspects of the environment that are spatially extensive. The challenge for cognitive neuroscience is to describe the visual capacities that are supported by this integration process and to discover how they are implemented in the brain.

Experimental manipulations that preserve the distribution of local stimulation while modulating the global percept can be used to investigate global integration (Sasaki, 2007). This often involves identifying and isolating aspects of local stimulation that are considered to be candidates for global integration. For example, the integration of local edges into

global shapes can be probed by using a spatial array of oriented elements in which the global arrangement of edge orientations either do or do not cohere into a perception of global form (Altmann, Bülthoff, & Kourtzi, 2003; Kourtzi, Tolias, Altmann, Augath, & Logothetis, 2003; Mannion, Kersten, & Olman, 2013). However, a potentially fruitful complementary strategy is to assess global integration in naturalistic environments, which contain a complex and rich spatial structure (Simoncelli & Olshausen, 2001) with myriad components that could potentially be targeted by global processes. Rather than assessing isolated cues, this strategy seeks to evaluate what cortical machinery is enabled, what processing pathways are traversed, and what visual capacities are brought online by the global structure of natural sensory stimulation.

Here, we used natural image patches to examine the sensitivity of low-level and mid-level human visual cortex to the global coherence of naturalistic local sensory stimulation. Observers viewed image patches through two circular apertures on the horizontal meridian in the visual field on either side of fixation (see Figure 1). We compared a globally coherent condition, in which the image patches were drawn from the same underlying extended image and evoked a compelling percept of global spatial structure, with a globally non-coherent condition, in which the image patches were drawn from different underlying extended images (see Onat, Jancke, & König, 2013, for a similar approach with natural movies). Critically, the distribution of local image patches within each aperture were identical for the two conditions over the duration of the experiment, which isolated sensitivity to global integration from variations in local sensory stimulation. Using functional magnetic resonance imaging (fMRI), we estimated the blood-oxygen level-dependent (BOLD) signal while human observers viewed such coherent and non-coherent stimuli to examine the consequences for the amplitude and spatial pattern of responses in human visual cortex.

Methods

Participants

Five observers (three female), each with normal or corrected-to-normal vision, participated in the current study. Each participant gave their informed written consent and the study conformed to safety guidelines for MRI research and was approved by the Institutional Review Board at the University of Minnesota.

Apparatus

Functional imaging was conducted using a 7T magnet (Magnex Scientific, UK) with a Siemens (Erlangen, Germany) console and head gradient set (Avanto). Images were collected with a T_2^* sensitive gradient echo imaging pulse sequence (TR = 2s, TE = 18ms, flip angle = 70°, matrix = 108 × 108, GRAPPA acceleration factor = 2, FOV= 162 × 162mm, partial Fourier= $\frac{7}{8}$, voxel size = 1.5mm isotropic) in 36 ascending interleaved coronal slices positioned such that the coverage extended slightly beyond the posterior end of the brain.

Stimuli were displayed on a screen positioned within the scanner bore using a VPL-PX10 projector (Sony, Tokyo, Japan) with a spatial resolution of 1024×768 pixels, temporal resolution of 60Hz, mean luminance of $168cd = m^2$, and an approximately linear relationship between video signal and projected luminance. Participants viewed the screen from a distance of 72cm, via a mirror mounted on the head coil, giving a viewing angle of $29.1° \times 21.8°$ that accommodated a visible square region of approximately 14.5° in length due to occlusion from the scanner bore. Stimuli were presented using PsychoPy 1.73.05 (Peirce, 2007). Behavioural responses were indicated via a FIU-005 fiber optic response device (Current Designs, PA). As detailed below, analyses were performed using FreeSurfer 5.1.0 (Dale, Fischl, & Sereno, 1999; Fischl, Sereno, & Dale, 1999), FSL 4.1.6 (Smith et al., 2004), and AFNI/SUMA (2013/09/20; R. W. Cox, 1996; Saad, Reynolds, Argall, Japee, & Cox, 2004). Experiment and analysis code is available at https://bitbucket.org/djmannion/ns_aperture.

Stimuli

The stimulus consisted of two circular apertures, presented on the horizontal meridian in the visual field on either side of fixation. Each aperture was 4° visual angle in diameter and was centred at 3° visual angle eccentricity, resulting in an nearest-edge horizontal distance of 2° visual angle. A circle of 0.15° visual angle in diameter was continually present at the centre of the display as a fixation and task indicator, and the remainder of the display was set to mid-grey (mean luminance). An illustration of the stimulus geometry is shown in Figure 1.

The images presented within the apertures were obtained from a publicly available natural image database (van Hateren & van der Schaaf, 1998). Each 1534×1024 pixel image was cropped to square regions, 140 pixels in length, corresponding to the location of the apertures. Each region was then normalised, separately, by subtracting its mean intensity and dividing by its maximum absolute intensity.

Images from the database were selected for inclusion in the study based on evaluation by the first author. A total of 108 images were selected, based on the subjective criterion that a compelling sense of globally coherent structure was evident when displayed with the limited field of view of the aperture geometry.

Design

Each experiment scanning run consisted of the blocked presentation of 216 events, with each event consisting of 1s stimulus display followed by $\frac{1}{3}$ s blank. The events were equally split into coherent and non-coherent experiment conditions, with each containing the full ensemble of 108 images. The event sequences were determined by either jointly (coherent) or separately (non-coherent) shuffling the presentation order of each aperture's image patches. By this procedure, each trial in the coherent condition consisted of image patches in the left and right apertures that were drawn from the same image, while the patches in the non-coherent condition were drawn from different images (see Figure 1 for an example). Events were ordered in 16s blocks (12 events) per condition, alternating between coherent and non-coherent blocks, with a total of 18 blocks per experiment run for an overall duration

of 288s. Each participant completed 10 such runs in a single session, with the starting block (coherent or non-coherent) alternating across runs.

Participants performed a behavioural rating task during each experiment run. On certain trials, at intervals drawn from a geometric distribution with a probability of 0.35, participants were cued via a change in the colour of the fixation marker to make a judgement of the current stimulus coherence on a four-point scale (confident coherent, less confident coherent, less confident non-coherent, confident non-coherent). The cue appeared 0.8s after stimulus onset to encourage observers to internally perform the judgement on each image presentation regardless of whether they received a subsequent cue to respond. Participants used different hands to make the coherent and non-coherent choices, with the particular hand assignment randomised at the beginning of each run.

Each participant also completed two runs, in the same session as the experiment runs, to localise the retinotopic location of the stimulus apertures. In alternating 16s blocks, interleaved with 16s blank screen baseline blocks, either the left or the right stimulus aperture was filled with a contrast-reversing (2Hz) checkerboard (2 cycles per degree). There were six such cycles per localiser run, prepended with an additional blank block of 22s duration, for a total duration of 310s.

Anatomical acquisition and processing

A T_1-weighted anatomical image (sagittal MP-RAGE, 1mm isotropic resolution) was collected from each participant in a separate session using a Siemens Trio 3T magnet (Erlangen, Germany). FreeSurfer (Dale et al., 1999; Fischl, Sereno, & Dale, 1999) was used for segmentation and cortical surface reconstruction of each participant's anatomical image, and to warp the resulting cortical surface into correspondence with FreeSurfer's standard surface template (Fischl, Sereno, Tootell, & Dale, 1999). SUMA was then used to convert the warped surfaces to a standard mesh (Saad et al., 2004).

Visual area localisation

Conventional retinotopic mapping and visual area localisation acquisition and analysis procedures, implemented as detailed in Mannion et al. (2013), were performed on each participant's standardised surface space. These surface datasets were combined across participants at each node on the surface, and the resulting maps of angular and eccentric visual field preference were used to assign likely visual area labels to low and mid-level visual cortex, as shown in Figure 2, to provide a framework for interpreting the location of regional activation in the main experiment. Standard criteria were used to delineate the borders of the low-level visual areas V1, V2, and V3 (Dougherty et al., 2003; Schira, Tyler, Breakspear, & Spehar, 2009). The ventral mid-level human V4 region (hV4) was defined as a full contralateral hemifield representation extending posterior to the ventral V3 border (Arcaro, McMains, Singer, & Kastner, 2009; Goddard, Mannion, McDonald, Solomon, & Clifford, 2011; Wade, Brewer, Rieger, & Wandell, 2002). We delineated three regions of dorsal mid-level cortex; LO1, LO2, and V3A/B. The LO1 and LO2 regions were defined as two contralateral hemifield representations parallel to the dorsal V3 border and extending from the central fovea and stopping before the border of V3A/B (Larsson & Heeger, 2006).

Visual areas V3A (Tootell et al., 1997) and V3B (Press, Brewer, Dougherty, Wade, & Wandell, 2001) are difficult to distinguish, and we defined V3A/B as a combined area with a contralateral hemifield representation. The V3A/B area proceeded adjacent to peripheral V3 before extending anteriorly from a characteristic junction to run perpendicular to the peripheral extent of LO1/2 (Larsson & Heeger, 2006; Press et al., 2001).

Conventional motion and object functional localisers were also acquired for each participant to provide additional landmarks for interpreting locations on the cortical surface. General linear model (GLM) analyses were performed for the motion and object localisers for each participant on a standardised surface, and the beta estimates for the localising contrasts (motion versus static, intact versus scrambled) were entered into a one-sample t-test across participants. The resulting maps of statistical significance were thresholded at a liberal level ($p < 0.01$, one-tailed, uncorrected) and are shown in Figure 3 (upper and middle panels). A similar analysis was performed on the within-session aperture localisers, and is shown in Figure 3 (lower panels).

Pre-processing

Estimates of participant motion were obtained using AFNI, with reference to the volume acquired closest in time to a within-session fieldmap image, and were combined with unwarping parameters (obtained via FSL) before resampling with sinc interpolation. The participant's anatomical image was then coregistered with a mean of all the functional images via AFNI's `align_epi_anat.py`, using a local Pearson correlation cost function (Saad et al., 2009) and six free parameters (three translation, three rotation). Coarse registration parameters were determined manually and passed to the registration routine to provide initial estimates and to constrain the range of reasonable transformation parameter values. The motion-corrected and unwarped functional data were then projected onto a standardised cortical surface by averaging the volume data between the white matter and pial boundaries (identified with FreeSurfer) using AFNI/SUMA. For the univariate analysis, surface-based spatial smoothing was performed on each run's timeseries using SUMA's `SurfSmooth`, calculated along a surface intermediate to the white matter and pial surfaces, to a full-width at half-maximum of 2.5mm. All analyses were performed on the nodes of this standardised surface domain representation.

Analysis

Univariate—First-level (participant-level) univariate analysis was conducted within a GLM framework using AFNI. Timecourses corresponding to coherent stimulus condition blocks were convolved with SPM's canonical haemodynamic response function and entered as a regressor in the GLM design matrix. Legendre polynomials up to the second degree were included as additional regressors. The first and last blocks of each run were censored in the analysis, leaving 1280 data timepoints (128 per run for 10 runs) and 31 regressors (1 stimulus and 30 polynomial) in the design matrix. The GLM was estimated via AFNI's `3dREMLfit`, which accounts for noise temporal correlations via an ARMA(1,1) model.

Second-level (group-level) statistical significance of the effect of coherent stimulation was assessed via a one-sample t-test on the beta weight assigned to the coherent stimulus

regressor in each participant's GLM (see Figure 5 for a representation of the single-participant beta weights). The *t*-test was performed against a null hypothesis of zero beta amplitude, and was conducted for all surface nodes for which acquisition coverage was achieved for all participants. To compensate for performing multiple comparisons (one comparison at each surface node within the acquisition region), we used a two-step procedure in which a height threshold of $p < 0.01$ (uncorrected) was followed by a cluster threshold of $p < 0.05$ (hemisphere family-wise error corrected). The cluster threshold was determined using AFNI's `slow_surf_clustsim.py` in a procedure which determines the distribution of cluster sizes obtained by applying the above analysis to 1000 random noise volumes. This produced cluster area thresholds of 147mm^2 and 176mm^2 for the left and right hemispheres, respectively.

Multivariate—The timeseries for each participant and run, projected onto a standard surface, were first high-pass filtered with Legendre polynomials up to the second degree. An amplitude was then estimated for each block (excluding the first and last blocks in each run) as the mean signal within its eight volumes (16s), shifted by three volumes (6s) to compensate for the delayed haemodynamic response. The amplitude estimates within each run were then normalized (z-scored). This procedure produced 160 responses per participant for each node on the cortical surface; 8 responses for each condition in each of 10 runs.

The MVPA was performed using a searchlight procedure (Kriegeskorte, Goebel, & Bandettini, 2006), in which a given surface node was designated, in turn, as a seed node and considered along with other nodes within a radius of 5mm along the cortical surface (midway between the white matter and pial surfaces) to form the multivariate data pattern. The analysis was implemented using a 10-fold leave-one-run-out strategy in which the responses from a given run were designated, in turn, to form a 'test' set and the remaining runs to form the 'training' set. Each training set thus consisted of 154 examples, with coherent and non-coherent conditions equally represented. In each analysis fold, a linear support vector machine (SVM) was constructed on the labelled training set and the classification accuracy assessed on the data from the test set. Support vector machines were implemented with `svmlight` (Joachims, 1998), and the accuracy of each seed node was taken as the average correction classification over the 10 folds.

The group-level statistical significance of the classification performance of each seed node was assessed via a one-sample *t*-test against a chance performance level of 50%. The single-participant classification accuracy surfaces were spatially smoothed, with the same parameters as for the univariate analysis, prior to the group *t*-test. A comparable multiple comparisons control strategy to the univariate analysis was adopted, in which a height threshold of $p < 0.01$ (one-tailed) was applied followed by a cluster-level correction ($p < 0.05$) using the same parameters as for the univariate analysis.

Results

We first evaluated whether the BOLD response across the sampled area of human visual cortex was significantly modulated by the presence of coherent versus non-coherent image patches. A univariate GLM analysis revealed bilateral clusters of significantly elevated

activity ($p < 0.01$ height threshold, uncorrected, followed by $p < 0.05$ cluster threshold, FWE corrected) within dorsal regions of mid-level visual cortex, as shown in Figure 4 (upper panels). The most prominent activation was observed in the vicinity of the retinotopic regions LO1/2 and extending dorsally toward (and beyond, in the right hemisphere) the boundary of visual area V3A/B. The peak accuracy in this region is consistent with a location associated with the transverse occipital sulcus (TOS), which is ventral to the V3A/B representation of the lower vertical meridian (Nasr et al., 2011). Significant activation was also present in dorsal V3 and, in the right hemisphere, slightly into dorsal V2. However, such apparent dorsal V2 and V3 activity may be spillover from neighbouring areas, particularly given the lack of an expected counterpart activation cluster in ventral V2 and V3. When this group-level univariate analysis was assessed at the single-participant level, variation in the spatial profile of activity differences across participants was observed but the outcomes of the group-level analysis were qualitatively present across participants (as shown in Figure 5).

We then investigated whether the local spatial distribution of BOLD activity across the cortical surface contained information that could be used to discriminate the observation of coherent and non-coherent image patches. We used multivariate pattern analysis (MVPA) techniques (Haynes & Rees, 2006; Norman, Polyn, Detre, & Haxby, 2006) to quantify the representational content of small searchlight (Kriegeskorte et al., 2006) disks (10mm diameter) centred at each node on the cortical surface. As shown in Figure 4 (lower panels), this analysis revealed extensive regions of mid-level visual cortex with activity patterns capable of distinguishing the coherent and non-coherent stimulus conditions at levels significantly greater than chance ($p < 0.01$ height threshold, uncorrected, followed by $p < 0.05$ cluster threshold, FWE corrected). As with the univariate analysis, the single-participant profiles of MVPA accuracy (not shown) varied across participants but were qualitatively similar to the outcomes of the group-level MVPA analysis.

The locations of regions with significantly above chance classification accuracy are similar in the left and right hemispheres, and we describe prominent features of interest located relative to the borders of nearby retinotopic visual areas and relative to regions activated by functional localisers. We begin considering the spatial distribution of significant classification accuracy at the central foveal representation (see the lower panels of Figure 2 for the map of eccentricity preference) and moving dorsally, where we first observe significant classification accuracy within and near the retinotopic areas LO1 and LO2 and the TOS region—consistent with the results of the univariate analysis. The significant levels of accuracy extend dorsally into area V3A/B and into areas in the intraparietal sulcus (Swisher, Halko, Merabet, McMains, & Somers, 2007) and in anterior and dorsal direction beyond the far boundary of LO2. This latter cluster appears to be more dorsal than the human motion complex (Amano, Wandell, & Dumoulin, 2009; Huk, Dougherty, & Heeger, 2002; Kolster, Peeters, & Orban, 2010), and may be an anterior region of V3B. Moving to posterior dorsal cortex, we observe significant levels of classification accuracy in a region beyond the far eccentricity boundaries of dorsal V2 and V3 in low-level visual cortex, which is likely to be associated with the retrosplenial cortex (RSC; Nasr et al., 2011).

Beginning again at the central foveal representation, moving ventrally we observe a cluster of significant accuracy within hV4. Bilaterally, this cluster appears to be situated in a somewhat foveal-preferring region of hV4, with an additional hV4 cluster at mid-eccentricity preference present only in the right hemisphere. Both hemispheres show clusters of significant accuracy in ventral regions beyond the far eccentricities of hV4 in putative VO1 areas (Arcaro et al., 2009; Brewer, Liu, Wade, & Wandell, 2005), before reaching the extent of the brain coverage of our functional acquisitions. There are also bilateral accuracy clusters in posterior ventral cortex, beyond the posterior border of hV4, which we tentatively assign to the putative human posterior inferior temporal cluster of retinotopic regions identified by Kolster et al. (2010). These clusters lie within regions associated with high levels of category selectivity (Malach, Levy, & Hasson, 2002), and partially overlap with functionally localised regions preferring intact relative to scrambled objects (see Figure 3, middle panels).

To evaluate the likelihood of such univariate and multivariate results being caused by unequal attentional allocation to the two conditions, we analysed participants' responses on the during-scanning behavioural task, in which they judged whether the image patches were part of a coherent or non-coherent image and whether they were confident or less confident of their judgement. When the patches were coherent, participants responded coherent/confident on 59.14% (SE= 6.74%), coherent/less confident on 28.74% (SE= 4.56%), non-coherent/less confident on 6.85% (SE=1.84%), and non-coherent/confident on 5.27% (SE= 1.84%) of trials. Similarly, when the patches were non-coherent, participants responded non-coherent/confident on 66.37% (SE= 4.77%), non-coherent/less confident on 25.33% (SE= 5.65%), coherent/less confident on 4.79% (SE=0.62%), and coherent/confident on 3.52% (SE= 1.26%) of trials. There was no statistically significant interaction between the response proportions and the stimulus condition ($F_{3,12} = 0.73$, $p \gg 0.05$). There was also no statistically significant difference in the response times for the coherent and non-coherent conditions ($F_{1,4} = 2.81$, $p = 0.17$), with participants responding with an average latency of 712ms (SE= 32ms) and 748ms (SE=25ms) for trials in the coherent and non-coherent conditions, respectively. These results suggest that participants' ability, confidence, and execution in classifying the two stimulus conditions was not appreciably different for coherent and non-coherent presentation.

Discussion

In this study, we were interested in characterising the response of the posterior regions of the human brain when observers viewed two natural image patches drawn either from the same full image or different full images. This stimulus manipulation caused the local patches to be integrated into a globally coherent percept or to be perceived as two non-coherent patches. We report an increased BOLD signal to coherent relative to non-coherent stimulation in the retinotopic regions LO1/2, which are likely to be also associated with the TOS region of dorsal cortex. We also find the presence of patch coherence to have widespread consequences for the local spatial pattern of BOLD signals in mid-level regions of visual cortex. This local spatial distribution of BOLD signals in dorsal regions, including LO1/2, TOS, V3A/B, RSC, and areas of the IPS, were informative of stimulus coherence, as were those in ventral regions including hV4 and VO1.

The most prominent perceptual consequence of the coherent stimulus condition, relative to the non-coherent condition, is the ability to recover the three-dimensional spatial structure, layout, and geometry of the scene depicted in the extended image. Accordingly, the region of posterior cortex with significantly elevated BOLD signal during the coherent stimulus condition, and high levels of pattern classification accuracy, resided in a location consistent with the TOS (Nasr et al., 2011)—an area of the brain implicated in the processing of visual scenes (Grill-Spector, 2003; Hasson, Harel, Levy, & Malach, 2003). Our association of such activation with the TOS was based primarily on its positioning relative to the borders of retinotopic visual areas (Nasr et al., 2011) rather than from a scene-network functional localiser. However, the one participant for which we had collected such a functional localiser for an unrelated study provides additional support for our association of the activation with the TOS, with the approximate location of the TOS cluster (identified from a scenes versus faces and houses contrast) tending to overlap with our interpretation of the position of the TOS (see row P4 in Figure 5).

The role of the TOS, alternatively referred to as the dorsal scene responsive area (DS; Nasr et al., 2011) and the occipital place area (OPA; Dilks, Julian, Paunov, & Kanwisher, 2013), in scene processing remains unclear. Furthermore, the precise location of TOS—and its relationship with nearby or underlying retinotopic regions—are uncertain. However, it appears to be a critical node in the scene processing network, as disruption of TOS using transcranial magnetic stimulation causes a selective impairment in the ability to discriminate scenes (Dilks et al., 2013). Resting-state functional connectivity analysis shows the TOS to link with areas of the intraparietal sulcus, LO1/2, and object-selective cortex (Nasr, Devaney, & Tootell, 2013), and this connectivity may contribute to the significantly above chance classification accuracy observed amongst this network in this experiment. Overall, the results of the current study lend further support for a role of TOS in processing spatially extensive visual information that coheres into a globally interpretable scene.

The sense of spatial layout that accompanies the coherent stimulus condition may also underlie the ability to discriminate the coherent and non-coherent conditions at levels significantly greater than chance in the RSC. Of its many apparent roles (Vann, Aggleton, & Maguire, 2009), the RSC region, also known as the medial scene responsive area (MS; Nasr et al., 2011), has been particularly implicated in computations for navigation and environmental orientation (Epstein, 2008; Maguire, 2001). The RSC also strongly prefers familiar scenes over unfamiliar scenes (Epstein, Higgins, Jablonski, & Feiler, 2007), and this familiarity effect may underlie the RSC's appearance in the current paradigm; the shuffling procedure we adopted means that each non-coherent exemplar is likely to be not previously seen, whereas each of the coherent exemplars are observed once per run and thus may become familiar. The RSC is also recruited by tasks involving spatial judgements (Nasr et al., 2013), however we consider it unlikely that this role underlies the selectivity observed here as both coherent and non-coherent conditions involve the performance of a spatial task to judge patch coherency. Finally, we note that, together with the TOS and RSC, an area of the parrahippocampus denoted the parrahippocampal place area (PPA; Epstein & Kanwisher, 1998) is frequently nominated as a key area in scene and spatial layout processing. The coverage of our functional acquisitions did not include the PPA, however

we consider it likely the PPA region would be strongly activated by the coherent versus non-coherent comparison in the current study.

The coordination of the aperture patches also supports the recovery of spatially extensive surface and contour structures. This recovery is often accompanied by a sense of amodal perceptual completion, in which the apparent spatial structure of the underlying extended image is perceived to be present behind an occluding front surface. Given that such completion effects would likely be particularly evident in the fovea due to its positioning as the intervening territory between the two apertures, it is interesting that we observe a cluster with significantly above chance classification performance in a foveal region associated with hV4. Neurons in macaque V4 appear to modulate their activity when their receptive fields lie within an illusory surface (M. A. Cox et al., 2013), including amodal illusory surfaces. The above-chance classification performance observed in hV4 may also relate to the capacity for contour completion and complex feature selectivity afforded by the coherence between image patches, given that ventral regions of mid-level visual cortex, including visual area hV4, are sensitive to isolated global form (Mannion et al., 2013; Ostwald, Lam, Li, & Kourtzi, 2008; Wilkinson et al., 2000).

We did not observe significant modulation of activity or representational content in any of the low-level visual areas V1, V2, or V3 (given the caveat that the observed dorsal V2 and V3 selectivity appears epiphenomenanal). Under a purely feedforward view of information transmission through the cortical hierarchy, this insensitivity to coherence may be attributable to the smaller receptive fields of these areas (Amano et al., 2009; Winawer, Horiguchi, Sayres, Amano, & Wandell, 2010) preventing the stimulation in the two apertures from direct interaction. However, the abundant feedback and horizontal connectivity within visual cortex and the apparent utility in using higher-level knowledge to disambiguate lower-level processing in natural images (Bullier, 2001; Epshtein, Lifshitz, & Ullman, 2008; Olshausen & Field, 2005) render it somewhat surprising that no low-level effect of coherence was observed. Although it is always difficult to interpret the lack of an effect, we suggest three possible reasons why we did not observe significant differences in low-level areas in the current experiment. First, our use of an abrupt aperture edge may have obscured any effect of coherence on the spatial spread of cortical activation. Using a similar presentation paradigm, in which natural image movie sequences were presented within restricted apertures, Onat et al. (2013) found that coherent versus non-coherent stimulation affected the magnitude of cat area 18 activity and increased the spatial spread of activation along the cortical surface connecting the two apertures. With our abrupt, rather than smooth, edge on the apertures, any inevitable small eye movements, uncorrelated with the stimulus condition, would have introduced comparatively large effects at the aperture borders that may have limited the ability to detect the finer modulation in the extent of within-aperture activity in coherent versus non-coherent stimulus conditions. Second, our study have been insufficiently powered to detect any differences occurring at the level of low-level visual cortex. Third, our use of a temporally-blocked stimulus design, while appropriate for detecting gross differences between coherent and non-coherent activity, may not have been sensitive to non-feedforward processing in the current context. While non-feedforward processing may have differentially affected the low-level activity during coherent and non-coherent conditions, the precise consequences of this effect may have been specific to the

particular image. This specificity may have rendered an inconsistent net effect in the magnitude and, in particular, spatial pattern of activation within stimulus condition blocks.

The outcomes of this study are unable to support precise claims about the joint image properties that distinguish coherent from non-coherent pairings. This is an important, and challenging, question for future research; properties such as luminance, chromaticity, edge structure, motion, spatial scale, and many others—and their interactions—could serve as cues to global coordination. A potentially fruitful avenue for future research is to utilise the tendency for observers to occasionally perceive coherent pairings as non-coherent (and vice-versa), which offers a means of dissociating perceived coherence from the coordination evident in a particular image pairing. In addition, a more detailed analysis could be obtained by using a condition-rich design where the responses to many coherent and non-coherent image pairings are obtained, and methods such as representational similarity analysis (Kriegeskorte, Mur, & Bandettini, 2008) used to evaluate various candidate models of potential cues derived from analysis of the joint image statistics. In the interim, we hope that the outcomes and approach of the current study will be instructive for such future research.

In summary, we examined the implications of observing image patches from the same or different underlying extended image on the magnitude and spatial pattern of the fMRI BOLD response in human low and mid-level visual cortex. Our goal was to identify the brain regions in this visual network sensitive to the coordination of spatially disparate image structure. We find that bilateral areas within and near the TOS region in dorsal mid-level cortex increased in the magnitude of activity while participants observed coherent rather than non-coherent patch pairings. Furthermore, we find that extensive regions of mid-level cortex contained information that could discriminate the global coherence of the image patches at levels significantly greater than chance. These results demonstrate the capacity of processing pathways in the human visual system to globally integrate local naturalistic sensory stimulation, which provides a platform from which the functional properties of the identified regions of visual cortex can be further characterised and their role in natural visual perception elucidated.

Acknowledgments

References

Altmann CF, Bülthoff HH, Kourtzi Z. Perceptual organization of local elements into global shapes in the human visual cortex. Current Biology. 2003; 13(4):342–349. [PubMed: 12593802]

Amano K, Wandell BA, Dumoulin SO. Visual field maps, population receptive field sizes, and visual field coverage in the human MT+ complex. Journal of Neurophysiology. 2009; 102(5):2704–2718. [PubMed: 19587323]

Arcaro MJ, McMains SA, Singer BD, Kastner S. Retinotopic organization of human ventral visual cortex. Journal of Neuroscience. 2009; 29(34):10638–10652. [PubMed: 19710316]

Brewer AA, Liu J, Wade AR, Wandell BA. Visual field maps and stimulus selectivity in human ventral occipital cortex. Nature Neuroscience. 2005; 8(8):1102–1109.

Bullier J. Integrated model of visual processing. Brain Research Reviews. 2001; 36(2–3):96–107. [PubMed: 11690606]

Cox MA, Schmid MC, Peters AJ, Saunders RC, Leopold DA, Maier A. Receptive field focus of visual area V4 neurons determines responses to illusory surfaces. Proceedings of the National Academy of Sciences of the United States of America. 2013; 110(42):17095–17100. [PubMed: 24085849]

Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical Research. 1996; 29(3):162–173. [PubMed: 8812068]

Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis. I: Segmentation and surface reconstruction. Neuroimage. 1999; 9(2):179–194. [PubMed: 9931268]

Dilks DD, Julian JB, Paunov AM, Kanwisher N. The occipital place area is causally and selectively involved in scene perception. Journal of Neuroscience. 2013; 33(4):1331–1336. [PubMed: 23345209]

Dougherty RF, Koch VM, Brewer AA, Fischer B, Modersitzki J, Wandell BA. Visual field representations and locations of visual areas V1/2/3 in human visual cortex. Journal of Vision. 2003; 3:586–598. [PubMed: 14640882]

Epshtein B, Lifshitz I, Ullman S. Image interpretation by a single bottom-up top-down cycle. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105(38): 14298–14303. [PubMed: 18796607]

Epstein RA. Parahippocampal and retrosplenial contributions to human spatial navigation. Trends in Cognitive Sciences. 2008; 12(10):388–396. [PubMed: 18760955]

Epstein RA, Higgins JS, Jablonski K, Feiler AM. Visual scene processing in familiar and unfamiliar environments. Journal of Neurophysiology. 2007; 97(5):3670–3683. [PubMed: 17376855]

Epstein RA, Kanwisher N. A cortical representation of the local visual environment. Nature. 1998; 392(6676):598–601. [PubMed: 9560155]

Fischl B, Sereno MI, Dale AM. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. Neuroimage. 1999; 9(2):195–207. [PubMed: 9931269]

Fischl B, Sereno MI, Tootell RB, Dale AM. High-resolution intersubject averaging and a coordinate system for the cortical surface. Human Brain Mapping. 1999; 8(4):272–284. [PubMed: 10619420]

Goddard E, Mannion DJ, McDonald JS, Solomon SG, Clifford CWG. Color responsiveness argues against a dorsal component of human V4. Journal of Vision. 2011; 11(4):1–21.

Grill-Spector K. The neural basis of object perception. Current Opinion in Neurobiology. 2003; 13(2): 159–166. [PubMed: 12744968]

Hartline HK. The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. American Journal of Physiology. 1938; 121:400–415.

Hasson U, Harel M, Levy I, Malach R. Large-scale mirror-symmetry organization of human occipito-temporal object areas. Neuron. 2003; 37(6):1027–1041. [PubMed: 12670430]

Haynes J-D, Rees G. Decoding mental states from brain activity in humans. Nature Reviews Neuroscience. 2006; 7(7):523–534.

Huk AC, Dougherty RF, Heeger DJ. Retinotopy and functional subdivision of human areas MT and MST. Journal of Neuroscience. 2002; 22:7195–7205. [PubMed: 12177214]

Joachims, T. Advances in kernel methods: Support vector machines. Cambridge, MA: MIT Press; 1998. Making large-scale support vector machine learning practical.

Kolster H, Peeters R, Orban GA. The retinotopic organization of the human middle temporal area MT/V5 and its cortical neighbors. Journal of Neuroscience. 2010; 30(29):9801–9820. [PubMed: 20660263]

Kourtzi Z, Tolias AS, Altmann CF, Augath M, Logothetis NK. Integration of local features into global shapes: monkey and human fMRI studies. Neuron. 2003; 37(2):333–346. [PubMed: 12546827]

Kriegeskorte N, Goebel R, Bandettini P. Information-based functional brain mapping. Proceedings of the National Academy of Sciences of the United States of America. 2006; 103:3863–3868. [PubMed: 16537458]

Kriegeskorte N, Mur M, Bandettini P. Representational similarity analysis - connecting the branches of systems neuroscience. Front Syst Neurosci. 2008; 2:4. [PubMed: 19104670]

Larsson J, Heeger DJ. Two retinotopic visual areas in human lateral occipital cortex. Journal of Neuroscience. 2006; 26(51):13128–13142. [PubMed: 17182764]

Maguire EA. The retrosplenial contribution to human navigation: a review of lesion and neuroimaging findings. Scandinavian Journal of Psychology. 2001; 42(3):225–238. [PubMed: 11501737]

Malach R, Levy I, Hasson U. The topography of high-order human object areas. Trends in Cognitive Sciences. 2002; 6(4):176–184. [PubMed: 11912041]

Mannion DJ, Kersten DJ, Olman CA. Consequences of polar form coherence for fMRI responses in human visual cortex. Neuroimage. 2013; 78:152–158. [PubMed: 23608060]

Nasr S, Devaney KJ, Tootell RBH. Spatial encoding and underlying circuitry in scene-selective cortex. Neuroimage. 2013; 83:892–900. [PubMed: 23872156]

Nasr S, Liu N, Devaney KJ, Yue X, Rajimehr R, Ungerleider LG, Tootell RBH. Scene-selective cortical regions in human and nonhuman primates. Journal of Neuroscience. 2011; 31(39):13771–13785. [PubMed: 21957240]

Norman KA, Polyn SM, Detre GJ, Haxby JV. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends in Cognitive Sciences. 2006; 10(9):424–430. [PubMed: 16899397]

Olshausen BA, Field DJ. How close are we to understanding V1? Neural Computation. 2005; 17(8):1665–1699. [PubMed: 15969914]

Onat S, Jancke D, König P. Cortical long-range interactions embed statistical knowledge of natural sensory input: a voltage-sensitive dye imaging study. F1000Research. 2013; 2:51. [PubMed: 24358899]

Ostwald D, Lam JM, Li S, Kourtzi Z. Neural coding of global form in the human visual cortex. Journal of Neurophysiology. 2008; 99(5):2456–2469. [PubMed: 18322002]

Peirce JW. Psychopy–psychophysics software in python. Journal of Neuroscience Methods. 2007; 162:8–13. [PubMed: 17254636]

Press WA, Brewer AA, Dougherty RF, Wade AR, Wandell BA. Visual areas and spatial summation in human visual cortex. Vision Research. 2001; 41(10–11):1321–1332. [PubMed: 11322977]

Saad ZS, Glen DR, Chen G, Beauchamp MS, Desai R, Cox RW. A new method for improving functional-to-structural MRI alignment using local Pearson correlation. Neuroimage. 2009; 44(3):839–848. [PubMed: 18976717]

Saad, ZS.; Reynolds, RC.; Argall, B.; Japee, S.; Cox, RW. SUMA: an interface for surface-based intra- and inter-subject analysis with AFNI; Proc. IEEE int biomedical imaging: Nano to macro symp; 2004. p. 1510-1513.

Sasaki Y. Processing local signals into global patterns. Current Opinion in Neurobiology. 2007; 17(2):132–139. [PubMed: 17369036]

Schira MM, Tyler CW, Breakspear M, Spehar B. The foveal confluence in human visual cortex. Journal of Neuroscience. 2009; 29(28):9050–9058. [PubMed: 19605642]

Simoncelli EP, Olshausen BA. Natural image statistics and neural representation. Annual Review of Neuroscience. 2001; 24:1193–1216.

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, Matthews PM. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage. 2004; 23(Suppl 1):S208–S219. [PubMed: 15501092]

Swisher JD, Halko MA, Merabet LB, McMains SA, Somers DC. Visual topography of human intraparietal sulcus. Journal of Neuroscience. 2007; 27(20):5326–5337. [PubMed: 17507555]

Tootell RB, Mendola JD, Hadjikhani NK, Ledden PJ, Liu AK, Reppas JB, Dale AM. Functional analysis of V3A and related areas in human visual cortex. Journal of Neuroscience. 1997; 17(18):7060–7078. [PubMed: 9278542]

van Hateren JH, van der Schaaf A. Independent component filters of natural images compared with simple cells in primary visual cortex. Proceedings: Biological Sciences. 1998; 265(1394):359–366.

Vann SD, Aggleton JP, Maguire EA. What does the retrosplenial cortex do? Nature Reviews Neuroscience. 2009; 10(11):792–802.

Wade AR, Brewer AA, Rieger JW, Wandell BA. Functional measurements of human ventral occipital cortex: retinotopy and colour. Philosophical Transactions of the Royal Society of London Biological Sciences. 2002; 357(1424):963–973.

Wilkinson F, James TW, Wilson HR, Gati JS, Menon RS, Goodale MA. An fMRI study of the selective activation of human extrastriate form vision areas by radial and concentric gratings. Current Biology. 2000; 10(22):1455–1458. [PubMed: 11102809]

Winawer J, Horiguchi H, Sayres RA, Amano K, Wandell BA. Mapping hV4 and ventral occipital cortex: the venous eclipse. Journal of Vision. 2010; 10(5):1. [PubMed: 20616143]
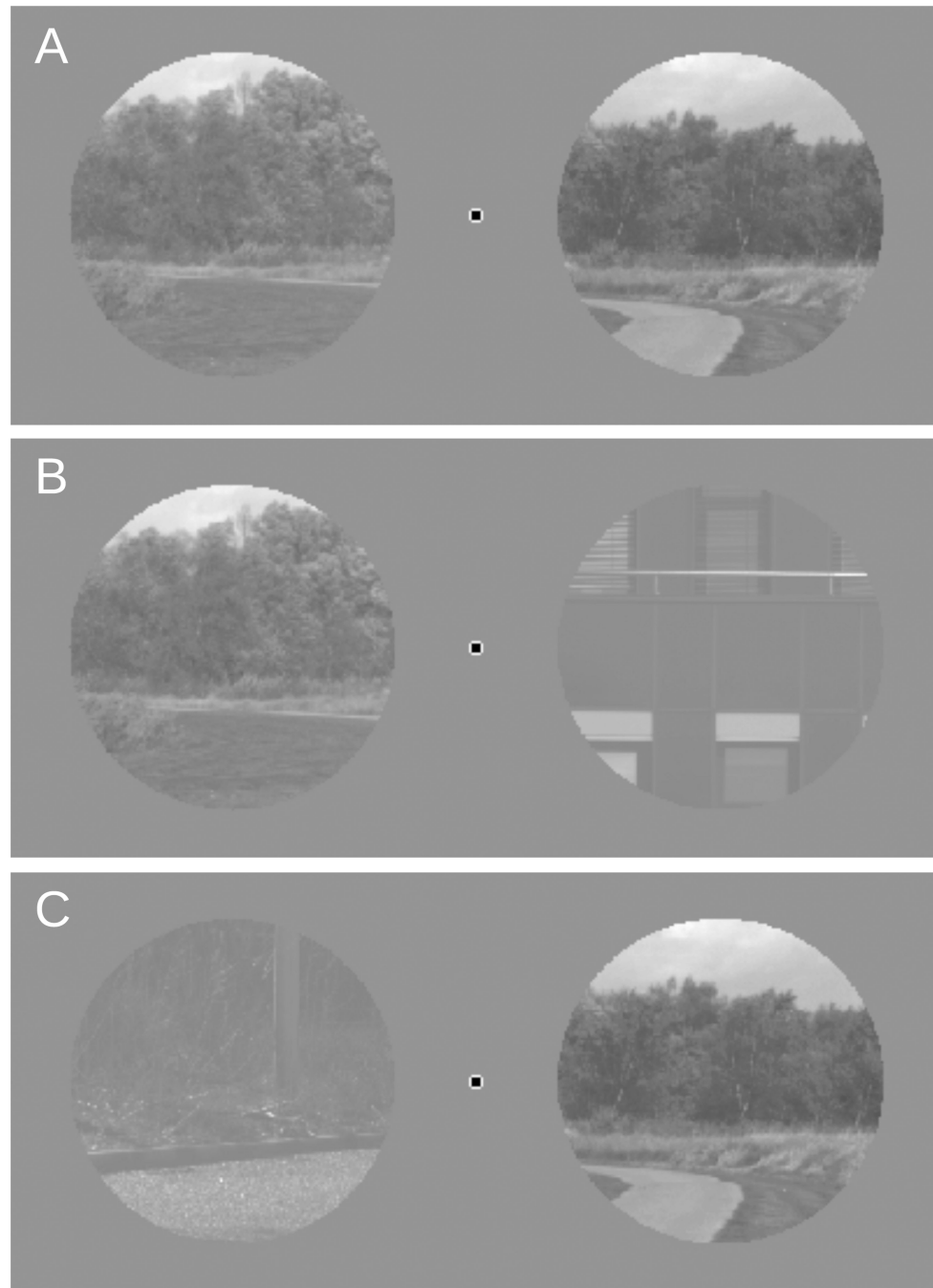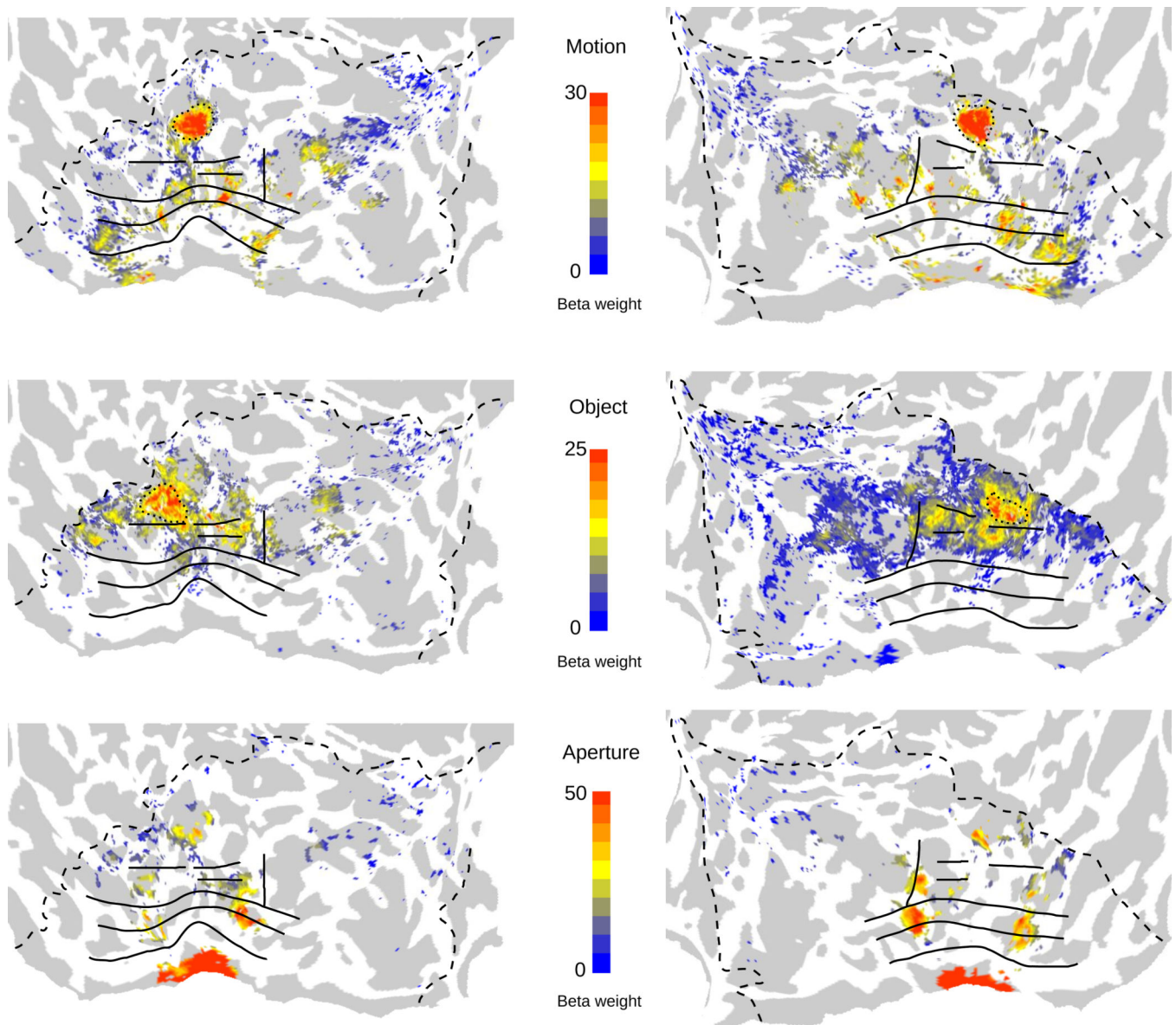
Figure 1.
Stimulus layout and conditions. Observers fixated a central marker and viewed two apertures, each 4° visual angle in diameter centred at 2° visual angle eccentricity, on the horizontal meridian. The apertures either showed local image regions from the same global image (panel A; coherent condition) or from different images (panels B and C; non-coherent condition).
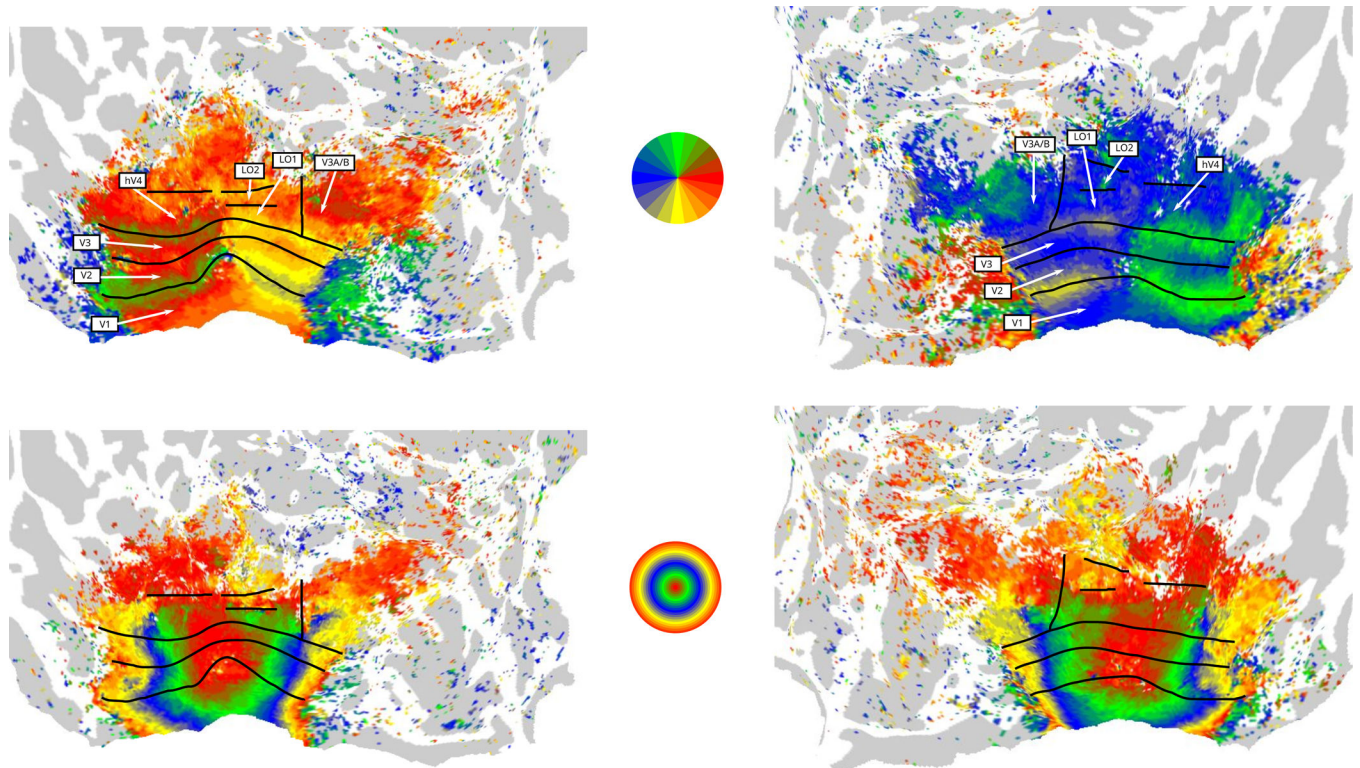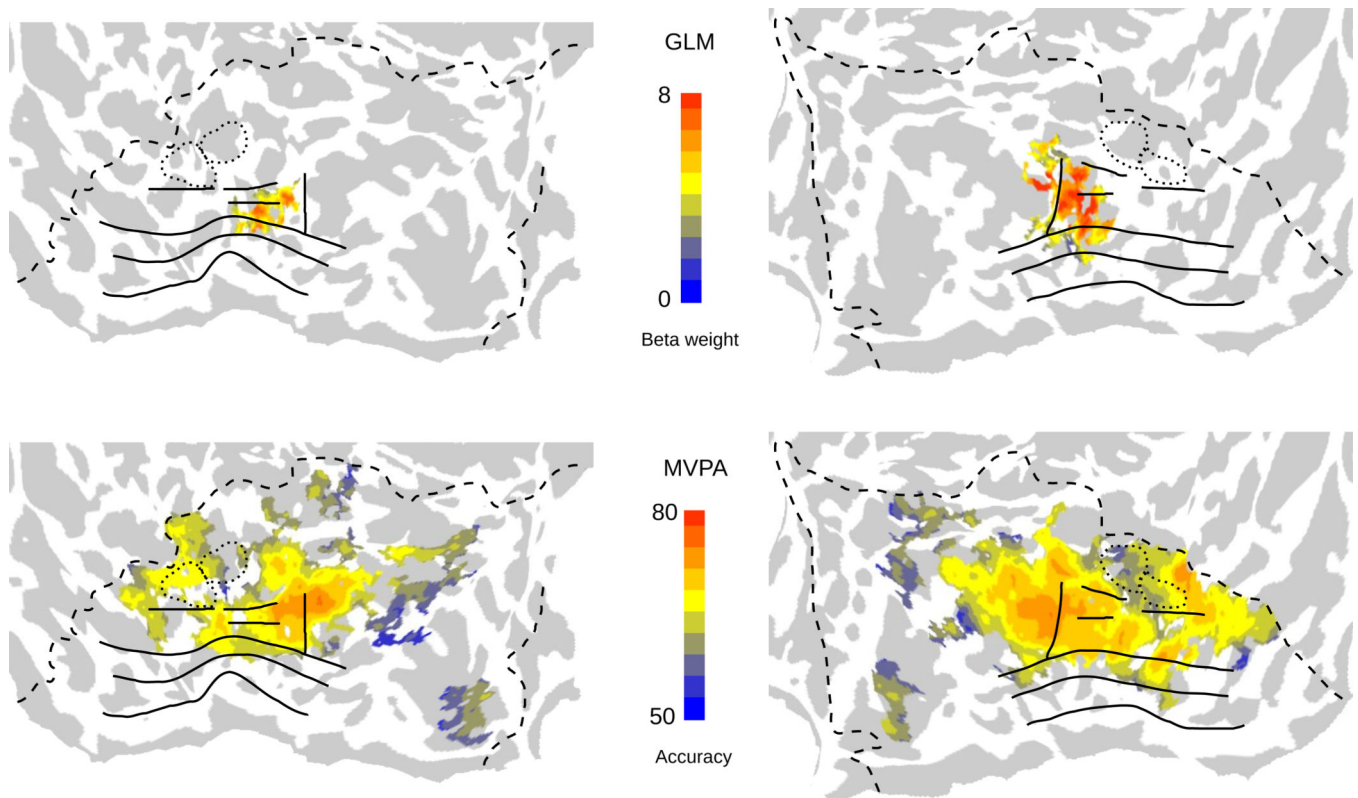
Figure 2.
Retinotopic visual area localisation. The upper and lower panels show the angle and
eccentricity, respectively, maps on a flattened representation of a group-average anatomical
left and right hemispheres (left and right columns, respectively). Boundaries for the low-
level visual areas V1, V2, and V3 and the mid-level visual areas LO1, LO2, V3A/B, and
hV4 are marked based on characteristic spatial progressions of the preferred visual field
position.

Figure 3.

Functional localisers. Panels show regions of the cortical surface with BOLD activity that was significantly elevated ($p < 0.01$, one-tailed, uncorrected) in functional localisation contrasts (upper: motion versus static, middle: intact versus scrambled, lower: left and right aperture checkerboards versus blank). Coarse boundaries around the most prominent clusters of interest for the motion and object localisers are shown as dotted lines. Unbroken dark lines indicate the borders of identified retinotopic visual areas, as per Figure 2, and dashed dark lines indicate the limits of the acquisition coverage in the experiment. Left and right columns show the left and right hemispheres, respectively.

Figure 4.
Univariate and multivariate group-level analysis results. The upper panels show the surface nodes with a significantly increased response to coherent versus non-coherent stimulus conditions, coloured according to the magnitude of the coherent stimulus regressor in the GLM analysis (arbitrary units). The lower panels show the surface nodes with significantly above chance performance in classifying coherent versus non-coherent stimulus conditions, coloured according to the level of accuracy (%). Significance for both univariate and multivariate results was determined by a height threshold of $p < 0.01$ (uncorrected) followed by a cluster threshold of $p < 0.05$ (FWE corrected). Each panel shows a flattened representation of the group-average brain, with dark lines showing the boundaries between the identified retinotopic visual areas (as per Figure 2), dotted lines enclosing the regions of most prominent activation to motion and object localisers (as per Figure 3), and dashed dark lines showing the boundary of the region from which functional signals were acquired. Left panels show the left hemisphere and right panels show the right hemisphere.
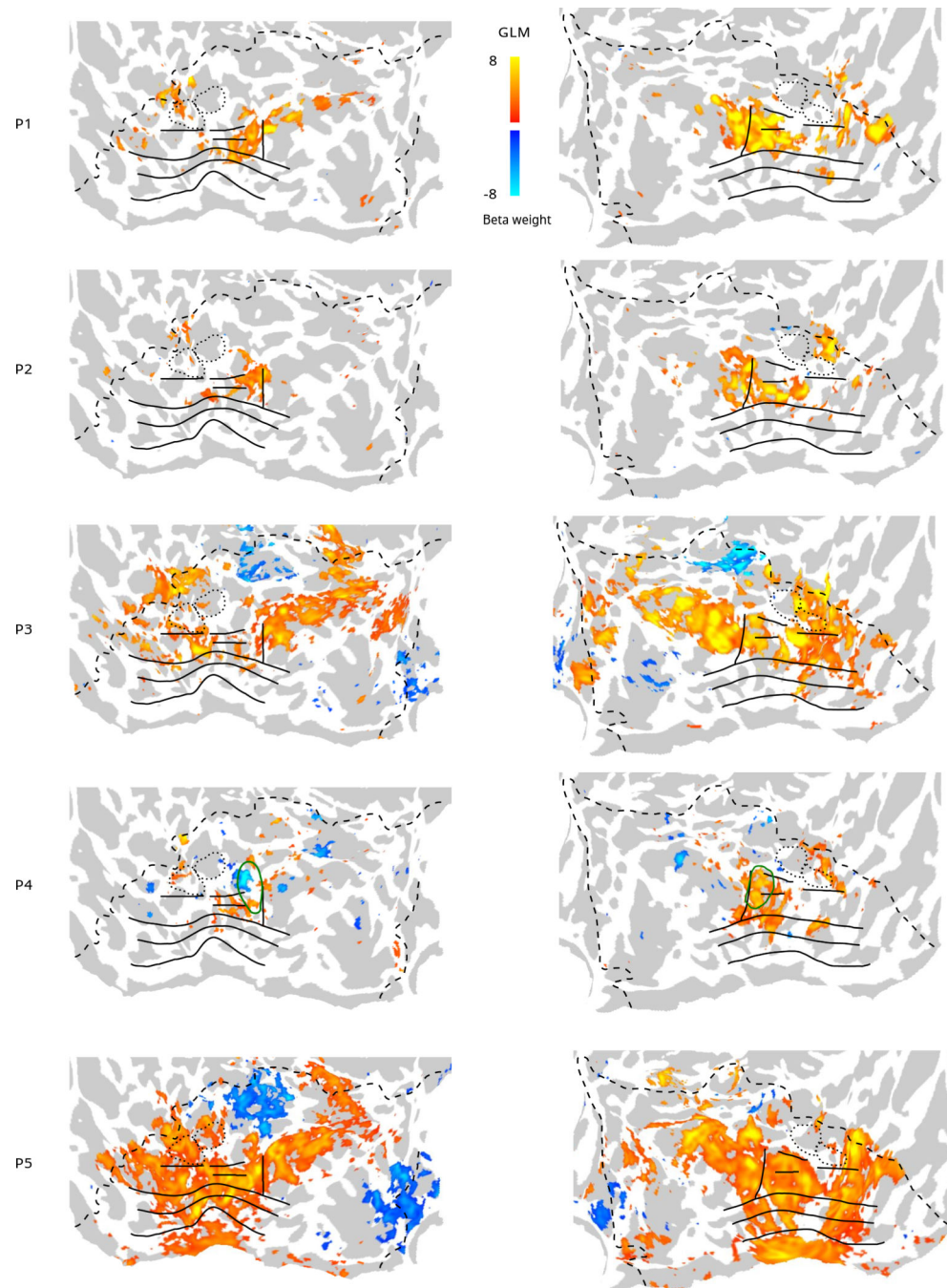
Figure 5.
Single-participant univariate results. Each panel shows the beta values for the coherent >
non-coherent contrast for each participant (rows) and hemisphere (columns). Values are
shown at a statistical significance threshold of $p < 0.001$ (uncorrected) on a flattened
representation of the group-average brain. Line markings are as per Figure 4, with the
exception of participant 4 (P4) for whom the green outline shows the boundaries of TOS as
estimated from a scene-network functional localiser. Note that the dashed lines represent the

extent of the acquisition coverage common to all participants, and hence may be exceeded at points by particular participants.