**Title:** Scene coherence can affect the local response to natural images in human V1

**Authors:**

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

**Affiliations:**

1. School of Psychology, UNSW Australia, Sydney, NSW, Australia
2. Department of Psychology, University of Minnesota, Minneapolis, MN, USA
3. Department of Brain and Cognitive Engineering, Korea University, Seoul, Korea

**Running head:** Scene coherence can affect V1 responses

**Contact information:**

Damien J. Mannion
School of Psychology
UNSW Australia, NSW, Australia
`d.mannion@unsw.edu.au`

**Counts:**

- Pages: 26
- Figures: 6
- Tables: 0
- Equations: 0
- Words (manuscript): 5685
- Words (abstract): 223
- Words (introduction): 672

**Keywords:**

context; visual cortex; fMRI; V1; surround suppression; scene perception

## Abstract

Neurons in primary visual cortex (V1) can be indirectly affected by visual stimulation positioned outside their receptive fields. Although this contextual modulation is intensely studied, we have little notion of how it manifests with naturalistic stimulation. Here, we investigated how the V1 response to a natural image fragment is affected by spatial context that is consistent or inconsistent with the scene from which it was extracted. Using fMRI at 7T, we measured the BOLD signal in human V1 (n=8) while participants viewed an array of apertures. Most apertures showed fragments from a single scene, yielding a dominant perceptual interpretation which participants were asked to categorize, and the remaining apertures each showed fragments drawn from a set of 20 scenes. We find that the V1 response was significantly increased for apertures showing image structure that was coherent with the dominant scene relative to the response to the same image structure when it was non-coherent. Additional analyses suggest that this effect was mostly evident for apertures in the periphery of the visual field, that it peaked towards the centre of the aperture, and that peaked in the middle to superficial regions of the cortical gray matter. These findings suggest that knowledge of typical spatial relationships is embedded in the circuitry of contextual modulation. Such mechanisms, possibly augmented by contributions from attentional factors, serve to increase the local V1 activity under conditions of contextual consistency.

Consider a small region of primary visual cortex (V1) in the occipital lobe of the human brain. The neurons within this region can be activated by visual stimulation in a particular part of the observer's visual field. However, the visual image in this receptive field is often insufficient for predicting the neural response properties—such responses can be affected by the visual stimulation present in surrounding parts of the visual field (Allman *et al.* 1985). Knowledge of the functioning of this contextual modulation is critical to understanding the role of V1 in the processing of visual information.

An important clue to the functional role of contextual modulation is that it can depend on the relationship of visual properties in the centre and surround. Perhaps the most well-studied example of this is orientation-dependent surround suppression, in which the V1 response to an oriented grating is typically reduced when surrounded by a grating of similar orientation compared to when surrounded by a grating of a dissimilar orientation (DeAngelis *et al.* 1994). As the orientation structure of center and surround regions in natural images are often similar (Felsen *et al.* 2005; Geisler *et al.* 2001), this dependence has been considered to reflect a computational strategy informed by the statistical structure of typical visual input (Schwartz *et al.* 2007).

Despite the apparent connection with the statistical properties of natural images, the effect of context on V1 responses has only rarely been assessed in the presence of naturalistic spatial structure. Vinje & Gallant (2000, 2002) showed that the activity of single neurons in macaque V1 is affected as the surrounding context changes from a uniform field to natural image structure. Onat *et al.* (2013) used voltage-sensitive dye imaging to show that cat primary visual cortex is affected by the presence of naturalistic spatial context. Importantly, Onat *et al.* (2013) also showed that the magnitude of contextual modulation depended on whether the context was drawn from the same natural scene—suggesting of a role for scene coherence in contextual modulation.

Here, our aim was to investigate the role of coherent spatial context in the modulation of population-level V1 responses in human visual cortex. Our primary hypothesis was that the degree of modulation in the response of a V1 region would be affected by the meaningful correspondence of its stimulation with that of its surrounding area—that is, whether the surround contains image structure that would be likely to be encountered given the stimulation in the central region.

In addition to the presence of contextual modulation with coherent natural stimulation, it is also important to consider the direction of the modulation—whether it evokes a increase or decrease in neural activity. Although often associated with suppression, contextual modulation can also have facilitatory effects on V1 activity (e.g. Gilad *et al.* 2013). In studies using natural stimuli, Vinje & Gallant (2000, 2002) showed that while the contextual effects at the single neuron level were mostly suppressive, facilitatory influences were also observed. Onat *et al.* (2013) found that coherent naturalistic context increased the population response compared to non-coherent spatial context. Because we investigated population-level responses and used a similar manipulation of surround coherence, we predicted that the response of human V1 would be most similar to that reported by Onat *et al.* (2013) and would show increased response levels to coherent naturalistic spatial context.

To test our hypotheses, we used functional magnetic resonance imaging (fMRI) to infer the magnitude of human V1 activity evoked by image structure that had either coherent or non-coherent surrounding context. We first identified the local V1 regions responsive to stimulation in a set of circular patches that tiled the visual field. We then quantified the response of such regions to stimuli in which the majority of patches depicted a single visual scene while the remainder were drawn from different scenes. This manipulation allowed us to compare the response of each V1 region under conditions in which its surrounding image structure was more likely or less likely to have been sourced from the same visual environment.

## Materials and Methods

### *Participants*

Nine observers (3 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. One participant was excluded from the analysis, as detailed below (§ Region of interest definition), and the presented results are based on the data from the remaining eight participants.

### *Apparatus*

Functional imaging was conducted using a 7T magnet (Magnex Scientific, UK) with a Siemens (Erlangen, Germany) console and head gradient set (AC84). Images for the main experiment and for the aperture localisation were collected with a $T_2^*$ sensitive gradient echo imaging pulse sequence (TR = 2s, TE = 20ms, flip angle = 70°, matrix = 128 × 128, GRAPPA acceleration factor = 2, FOV= 128 × 128mm, partial Fourier= $^6/_8$, voxel size = 1mm isotropic) in 36 ascending interleaved coronal slices covering the occipital lobes. For one participant's localiser session, a voxel size of 1.5mm isotropic was used (other parameters were as above except: matrix = 108 × 108, FOV=162 × 162mm, and 38 slices).

Stimuli were displayed on a screen positioned within the scanner bore using an NP4100 projector (NEC, Toyko, Japan) with a spatial resolution of 1024 × 768 pixels and temporal resolution of 60Hz. A gamma value of 2 was applied to the video card output to coarsely correct for its non-linear relationship with projector output. Participants viewed the screen from a distance of 72cm, via a mirror mounted on the head coil, giving a viewing angle of 27.5° × 20.7°. Stimuli were presented using PsychoPy 1.75.01 (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 *et al.* 1999; Fischl *et al.* 1999), FSL 4.1.6 (Smith *et al.* 2004), and AFNI/SUMA (2012/11/23; Cox 1996; Saad *et al.* 2004). Details on the implementation of the referenced AFNI/SUMA commands is documented at `http://afni.nimh.nih.gov/pub/dist/doc/program_help/`. Experiment and analysis code is available at `http://bitbucket.org/djmannion/ns_patches`.

### *Stimuli*

The stimulus consisted of an array of apertures that each revealed an image patch, as shown in Figure 1. A total of 32 apertures were placed at different polar angles in the visual field in four iso-eccentric rings surrounding central fixation. The ring closest to fixation was at 1.6° eccentricity, and the 6 apertures in this ring were each 1.1° in diameter. The next ring was at 3° eccentricity, and the 10 apertures in this ring were each 1.5° in diameter. The third ring was at 5.3° eccentricity, and the 12 apertures in this ring were each 2° in diameter. The final ring was at 8.4° eccentricity, and the 4 apertures in this ring were each 2.8° in diameter. The opacity of each aperture was modulated from 60% to 100% of its radius with a raised cosine profile; that is, an inner circular region with a radius that was 60% of the aperture radius was presented at full contrast and the remaining outer region was smoothly ramped to invisibility. Our rationale for using such an aperture window was to avoid the strong edge responses associated with an abrupt transition and to prevent inevitable small eye movements from having a disproportionally large effect at the aperture edges (see Mannion *et al.* 2014, for further discussion). The background of the display was set to mid-grey.

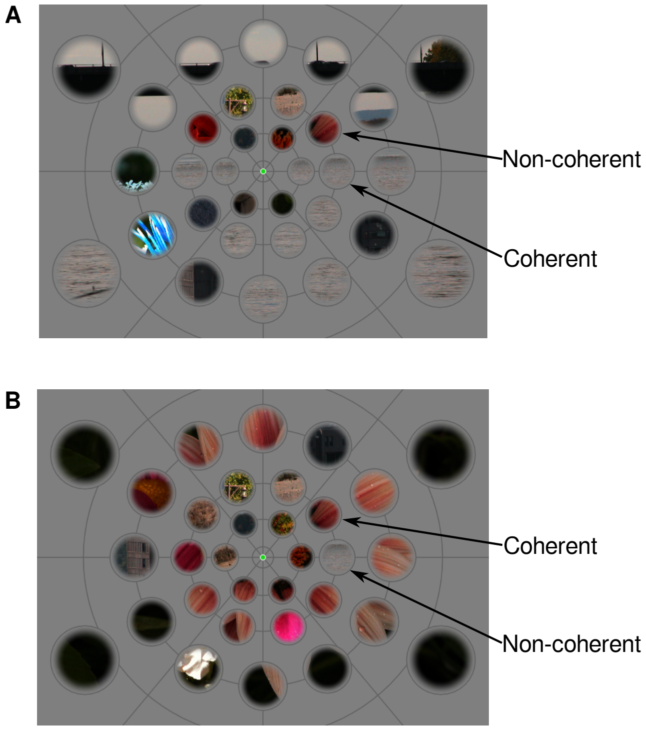


FIG. 1. Stimulus examples. Each stimulus is formed from an array of apertures that each shows an image patch; 19 of the 32 apertures show patches from the same image, while the remaining apertures show patches from images chosen pseudorandomly from others in the set. The stimulus thus has an overall dominant image—apertures that are consistent with this image are labelled as coherent while those that are inconsistent are labelled non-coherent. In panels **A** and **B**, the dominant image is a river and a plant with orange flowers, respectively. The two highlighted apertures are identical in panels **A** and **B** but vary in their coherence according to the dominant image.

The images presented within the apertures were obtained from the McGill Calibrated Colour Image Database (Olmos & Kingdom 2004). Each image was 768 × 576 and was linearised based on corrections for the non-linear response of the camera sensors (see Olmos & Kingdom 2004, and `http://tabby.vision.mcgill.ca` for specific details of this procedure). The image was then normalised to the maximum intensity that could be registered by the camera, which was obtained by running the linearisation and conversion procedure on a single pixel of maximum intensity.

Images from the database were selected for inclusion in the study based on evaluation by the first author. A total of 20 images were selected, based on the objective criterion that half (10) were to be primarily of flowers (to accommodate the behavioral task; see § Design) and the subjective criterion that a compelling sense of globally coherent structure was evident when displayed with the limited view of the aperture geometry.

### *Design*

The experiment followed a rapid event-related design protocol. Each scanning run consisted of 108 events, with an inter-event interval of 4 seconds, of which 100 events were generated in random sequence and the final 8 events were replicated and prepended. Of the 100 events, 20 were null events in which no stimulus was displayed and the remaining events involved presentation of the aperture display for the initial 2 seconds of the event.

For each of the 80 stimulus events, a given image was designated as the coherent image for that event; with 20 images in the set, each image was the coherent image on four events in each run. The four apertures in the furthermost eccentricity ring and the two apertures situated on the vertical meridian always displayed the patch from the coherent image, while 13 of the remaining 26 apertures were selected to show the patch from the coherent image and the remaining 13 apertures displayed patches from other images in the set (chosen pseudorandomly). The assignment of coherent and non-coherent patches was accomplished such that each patch displayed each image as part of the coherent stimulus on two events and as part of the non-coherent image on two events in each run. An example stimulus trial sequence for a particular aperture for a single run of the experiment is shown in Supplementary Figure 1. The overall duration of each run was 432 seconds, and participants completed 8–10 runs that were collected in a single session.

Participants performed a forced-choice categorisation task during each stimulus event, in which they were asked to respond via button press whether the coherent image shown was of flowers or non-flowers. This was not a difficult task, and performance was at or close to ceiling for each participant (who also received feedback, via proportion correct, at the end of each run). The task was included in order to motivate participants to consider each stimulus and to allow monitoring of participant engagement.

### *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 *et al.* 1999) was used for segmentation, cortical surface reconstruction, and surface inflation and flattening of each participant's anatomical image.

*Region of interest definition*

The V1 region of occipital cortex was defined based on analysis of functional acquisitions, obtained in a separate scanning session (using the same 7T scanner as in the main experiment), that followed standard procedures for the delineation of retinotopic regions in human visual cortex. Participants observed four runs of a clockwise/anti-clockwise rotating wedge stimulus and two runs of an expanding/contracting ring stimulus (DeYoe *et al.* 1996; Engel *et al.* 1997; Hansen *et al.* 2007; Larsson & Heeger 2006; Sereno *et al.* 1995; Schira *et al.* 2007), and the data were analysed via phase-encoding methods (Engel 2012) to establish visual field preferences over the cortical surface; see Mannion *et al.* (2013) for details. The angular and eccentricity phase maps were used to manually define each participant's V1.

Each participant was also scanned in a separate session (using the same 7T scanner as in the main experiment) to delineate the region of their V1 that corresponded to the retinotopic location of each aperture in the array. One participant's scanning session was repeated after the first failed to establish a robust localisation of the V1 aperture locations. Each run in such a localiser session was a rapid event-related design sequence with 83 events, with an inter-event interval of 4 seconds, of which 75 events were generated in random sequence and the final 8 events were replicated and prepended. Of the 75 events, 18 were null events in which no stimulus was displayed and the remaining events involved presentation of a contrast-reversing square-wave grating for the initial 1 second of the event. A separate random sequence was generated for each aperture, and each aperture had an equal probability of having a 2 second offset added to their sequence to reduce the number of activated apertures at a given point in the sequence. Six such sets of sequences were generated, and each was checked using AFNI's `3dDeconvolve` for interpretability in a general linear model (GLM) framework. All participants were shown the same set of sequences over the course of a six-run session.

Images from the aperture localiser session were pre-processed (as outlined below) and were then analysed within GLM framework using AFNI. The event onset sequence for each of the 32 apertures in the stimulus array was convolved with SPM's canonical haemodynamic response function and included as regressors in the GLM design matrix. Legendre polynomials up to the third degree were also included as additional regressors. The first 16 volumes (32s) of each run were censored in the analysis, leaving 900 data timepoints (150 per run for 6 runs) and 56 regressors (32 stimulus and 24 polynomial) in the design matrix. The GLM was estimated via AFNI's `3dREMLfit`, which accounts for noise temporal correlations via a voxelwise ARMA(1,1) model. This procedure produced a $t$ statistic for each of the 32 aperture regressors at each node on the cortical surface within V1, which were evaluated at a statistical significance level of $p < 0.001$ (uncorrected). A node was assigned as belonging to a particular aperture if it responded significantly to the aperture's corresponding regressor, and did not have a significant response to any of the other apertures' regressors. The cortical surface map of nodes with an assigned aperture were then subjected to a cluster area threshold of 5mm$^2$. This procedure produced cortical maps of aperture-specific responsiveness in agreement with the retinotopic organisation of V1, as shown for an example participant in Figure 2.

Each aperture was considered as being reliably localised if it contained a minimum of 10 nodes for each

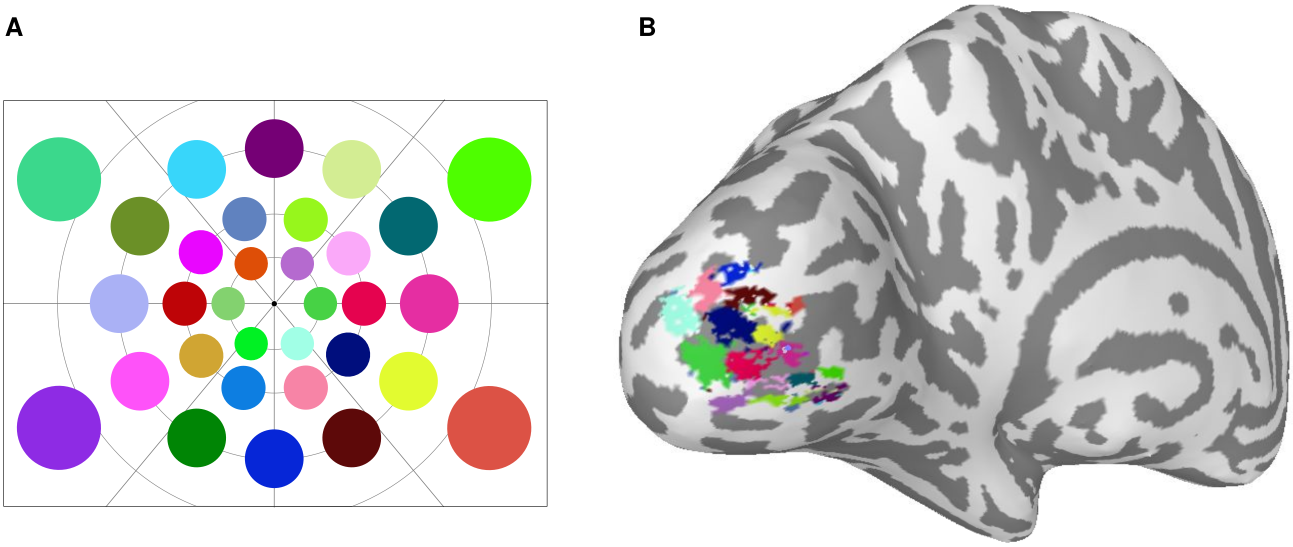


FIG. 2. Aperture geometry localisation in human V1. **A**: The position of each aperture in the displayed image, with each aperture assigned a different colour. **B**: The outcome of the localiser analysis for a single example participant, shown on an inflated view of their left hemisphere. Coloured areas of the cortical surface indicate V1 regions that were significantly ($p_{\text{uncorrected}} < 0.001$, 5mm$^2$ cluster threshold) modulated by the presence of image structure in one of the apertures, with the colour corresponding to the assignment in **A**.

participant. One participant had less reliable localisation overall, and was excluded from further analysis. After excluding this participant, 5 of the 26 apertures that modulated between coherent and non-coherent presentation were deemed to be not localised with sufficient reliability and were excluded from analysis in the main experiment. The number of nodes for each modulated aperture and each participant, including those excluded from further analysis, are presented in Supplementary Table 1.

*Pre-processing*

Images were corrected for differences in slice acquisition timing using AFNI, with reference to the first acquired slice. 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. The participant's anatomical image was then coregistered with a mean of all 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 the cortical surface by averaging along 15 evenly-spaced points between corresponding nodes on the (smoothed) white matter and pial surfaces (identified with FreeSurfer) using `3dVol2Surf` in AFNI/SUMA. Voxels situated within multiple points along a given white-pial segment were allowed to contribute multiple values to the average. No specific spatial smoothing was applied. All analysis was performed on the nodes of this surface domain representation in the participant's native brain space.

*Analysis*

Functional image analysis was conducted within a GLM framework using AFNI. We ran a separate GLM for each aperture that could be reliably localised (21 of the 26; see § Region of interest definition). For a given aperture, each of the 80 stimulus events in each run was assigned as either coherent or non-coherent (see § Design). These condition onsets were convolved with SPM's canonical haemodynamic response function and entered as regressors in the GLM design matrix. Legendre polynomials up to the third degree were included as additional regressors. The first 16 volumes (32s) of each run were censored in the analysis, leaving 200 data timepoints and 6 regressors (2 condition and 4 polynomial) per run in the design matrix. Additional polynomial regressors were added for each separate run, while the condition onsets were concatenated across runs to produce single regressors. Each GLM was estimated via AFNI's `3dREMLfit`.

The condition beta weights obtained from each GLM were converted to percent signal change (psc) via division by the average of the Legendre polynomial regressor timecourse at each node in the aperture. The psc values were then averaged across all nodes in the aperture and then across all apertures to yield an estimate of the response to coherent and non-coherent stimulus presentation conditions for each participant. These values were then normalised across participants by subtracting each participant's mean and adding the grand mean (Cousineau 2005). A paired-sample $t$ test was applied to investigate the hypothesis of response level differences between coherent and non-coherent conditions. Since the direction of this effect was uncertain, we applied a two-tailed test of significance to the outcome of this analysis.

We also conducted a series of additional exploratory analyses on the data from this study. For the first exploratory analysis, we investigated the effect of coherence separately for apertures at different eccentricities. As shown in Figure 1 and described in § Stimuli, each aperture was positioned in one of three rings that were each equidistant from central fixation (the four apertures in a fourth ring always displayed the coherent image and hence were not considered). We averaged the response to coherent and non-coherent conditions separately for apertures in the three rings.

We then examined whether an effect of coherence depended on the spatial location of the V1 representation within each aperture. We used SUMA's `SurfClust` to identify the centre node of each aperture for each participant (the centre node is the node for which the sum of distances to all other nodes in the aperture is minimum). We then used SUMA's `SurfDist` to calculate the minimum distance along the pial surface from a given aperture's centre node to all other nodes within the aperture (see Figure 5A). Internally, SUMA uses a standard implementation of Dijkstra's algorithm (Dijkstra 1959) to calculate such minimum distances between nodes. For each participant, we then averaged the response to coherent and non-coherent conditions for nodes based on their distance from the centre node, in 2mm wide bins with left edges equally spaced between 0 and 8mm and for nodes at distances greater than 10mm.

Finally, we investigated whether the coherence effect depended on the cortical depth within V1 (the relative distance between the white and pial surfaces). We defined a set of bins that were each 20% of the distance between the white and pial surfaces and placed at 20% intervals from 0% to 100% (Olman *et al.* 2012). We then used AFNI/SUMA's `3dVol2Surf` to average the timecourses of the voxels within each depth bin for each participant, forming a cortical surface representation for each participant, bin, and hemisphere. These surfaces were then analysed with the same GLM approach as applied for the main analysis, yielding an estimate of the response to coherent and non-coherent conditions at each depth bin.

## Results

We presented observers with natural image patches in an array of apertures that tiled the visual field. By altering the allocation of source images to the apertures, we manipulated the likelihood that a given aperture would be in the context of image structure from the same (coherent) or not from the same (non-coherent) scene. Importantly, a difference in the response to coherent and non-coherent presentations cannot be attributed to different local image properties—over the course of the experiment, each aperture displayed the same set of images in both the coherent and non-coherent conditions.

We find that coherent and non-coherent image patches evoked different levels of BOLD response in human V1, with coherent and non-coherent stimulation leading to an average of 1.60 percent signal change units (psc) and 1.51 psc respectively (normalised for differences in overall level of activation across participants; SEM= 0.01), as shown in Figure 3. This difference was statistically significant (paired sample $t_7 = 3.08, P = 0.018$). Hence, the local V1 response can be affected by the consistency of its surrounding context with the overall scene—with the response increasing for a coherent relative to a non-coherent context.

We conducted additional exploratory analyses to probe the characteristics of the apparent differences between the coherent and non-coherent conditions. First, we were interested in determining whether the coherence effect depended on aperture eccentricity. To investigate this, we calculated the response to coherent and non-coherent conditions separately for apertures in the three eccentricity rings in the array; inner, middle, and outer (see Figure 1). The magnitude of the coherent and non-coherent difference was significantly different across the eccentricity rings (interaction between coherence and eccentricity in a two-way repeated measures ANOVA; $F_{2,14} = 5.11, P = 0.022$). As shown in Figure 4, a significant difference between coherent and non-coherent conditions was evident in the middle and outer eccentricities but not at the inner eccentricity. For apertures at the inner eccentricity, coherent and non-coherent conditions evoked response magnitudes of 1.12 psc and 1.09 psc, respectively (paired sample $t_7 = 1.78, P = 0.118$). Response magnitudes were 1.82 psc and 1.71 psc for coherent and non-coherent conditions for the middle eccentricity apertures (paired sample $t_7 = 2.81, P = 0.026$), and 1.74 psc and 1.62 psc for apertures at the outer eccentricity (paired sample $t_7 = 3.47, P = 0.010$).

We then investigated whether the apparent difference between coherent and non-coherent stimulation depended on the position within each aperture. For each participant, we determined the centre of each aperture's V1

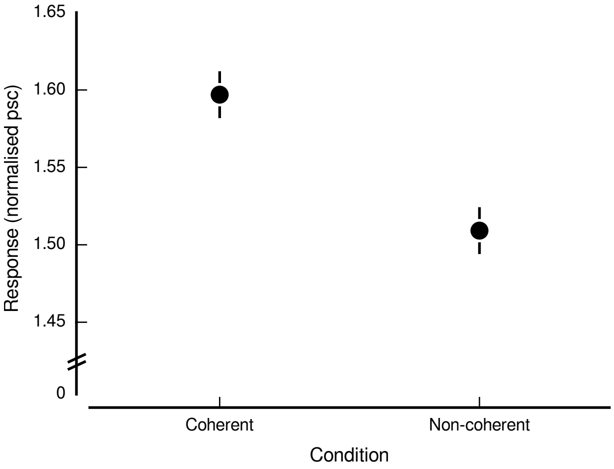


FIG. 3. Response in V1 to coherent and non-coherent image patches. The vertical axis shows the response amplitude (percent signal change units), and the horizontal axis shows the experiment conditions, with coherent and non-coherent depending on the relationship between an aperture's image patch and that of the other apertures in the display. The points show the BOLD response (normalised for differences in overall activation levels, across participants) averaged over participants, source images, and apertures, and the lines are ±1 SEM.

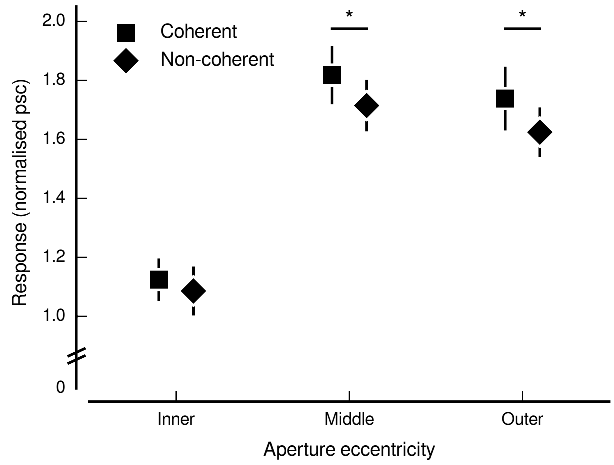


FIG. 4. Response in V1 to coherent and non-coherent image patches for apertures at different eccentricities. The vertical axis shows the response amplitude (percent signal change units), and the horizontal axis shows the eccentricity of the apertures. Points show the BOLD response (normalised for differences in overall activation levels, across participants) averaged over participants, source images, and apertures at a given eccentricity (squares and diamonds show coherent and non-coherent conditions, respectively), and the lines are ±1 SEM. Asterisks mark comparisons that are statistically significant ($p < 0.05$).

representation and then calculated the distance across the cortical surface of each aperture's constituent nodes (see Figure 5A). As shown in Figure 5B, the average BOLD response elicited by both coherent and non-coherent conditions decreased with distance from the aperture centre, reaching a minimum at approximately 8–10 mm from the centre. The magnitude of the difference between coherent and non-coherent responses was significantly different across the distances from the aperture centre (interaction between coherence and distance in a two-way repeated measures ANOVA; $F_{5,35} = 8.06, P < 0.001$), and displayed a significant negative linear trend (one-sample $t_7 = -4.78, P = 0.002$). As shown in Figure 5C, the difference between the responses to the coherent and non-coherent conditions was maximal close to the centre of the aperture and decreased until reaching parity at approximately 8–10 mm from the aperture centre.

Finally, we examined whether the difference between coherent and non-coherent stimulation may vary with cortical depth. By separately averaging voxels at different relative distances between the white matter and pial surfaces, we are able to get a coarse estimate of the distribution of activity at different relative cortical depths. As shown in Figure 6A, the response increases with increasing distance from the white matter for both coherent and non-coherent conditions; such a depth profile is consistent with the gradient-echo acquisition sequence used here (Zhao *et al.* 2004). There is a significant difference in the magnitude of coherent and non-coherent responses in each of the depth bins (all $P < 0.05$ based on paired-sample $t$-tests), as evident in Figure 6B. Notably, the magnitude of the difference is unequal across the bins (interaction between coherence and depth bin in a two-way repeated measures ANOVA; $F_{4,28} = 4.24, P = 0.008$)—a significant quadratic trend is evident with an inverse-U shape across depth bins (one-sample $t_7 = -2.99, P = 0.020$), with no other statistically significant polynomial trends (all $P > 0.05$ based on one-sample $t$-tests).

## Discussion

We investigated the response of human V1 to locally isolated patches of natural image structure. We were interested in whether such responses depend on whether the image structure was presented within a spatial context that was consistent or inconsistent with the image content—whether it was coherent or non-coherent with the overall visual impression. Using high resolution functional magnetic resonance imaging (fMRI), we found that the responses are affected by spatial context, with coherent presentation evoking significantly greater magnitudes of V1 activity than non-coherent presentation. Additional exploratory analyses suggest that this difference is most evident for image content away from central fixation, is localised within the spatial layout of the image structure, and is largest in the middle to upper regions of the cortical grey matter.

The findings of this study are consistent with the results of Onat *et al.* (2013), who used voltage-sensitive dye imaging to demonstrate that the local response in cat primary visual cortex to natural movies was facilitated by spatial context that was drawn from the same movie. The cortical depth profile we observed, in which the difference is greatest in the upper-middle section of the grey matter (where lowest is white matter and highest is pial surface), is also consistent with the likelihood that the optical imaging will have a greater contribution from superficial layers (Grinvald *et al.* 1999).

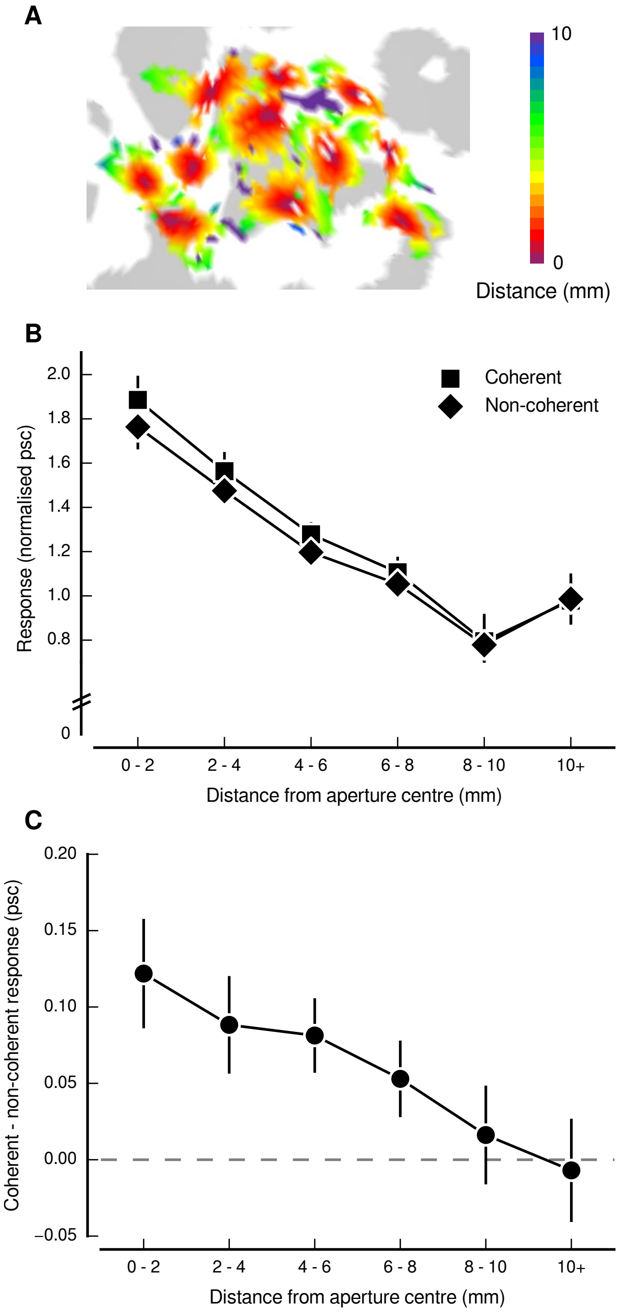


FIG. 5. Dependency of estimated response on the distance from aperture centre. **A**: An example participant right hemisphere (sphere view), showing the distance of each node associated with an aperture from the aperture centre. **B**: Response in V1 to coherent and non-coherent image patches for nodes at different distances from the aperture centre. The vertical axis shows the response amplitude (percent signal change units), and the horizontal axis shows the distance from the aperture centre (2mm bins). Points show the BOLD response (normalised for differences in overall activation levels, across participants) averaged over participants, source images, and nodes at a given distance from the aperture centre (squares and diamonds show coherent and non-coherent conditions, respectively), and the lines are ±1 SEM. **C**: Difference between the response to coherent and non-coherent image patches for nodes at different distances from the aperture centre. The vertical axis shows the difference amplitude (percent signal change units), and the horizontal axis shows the distance from the aperture centre (2mm bins). Points show the BOLD response difference (coherent − non-coherent) averaged over participants, source images, and nodes at a given distance from the aperture centre, and the lines are ±1 SEM.

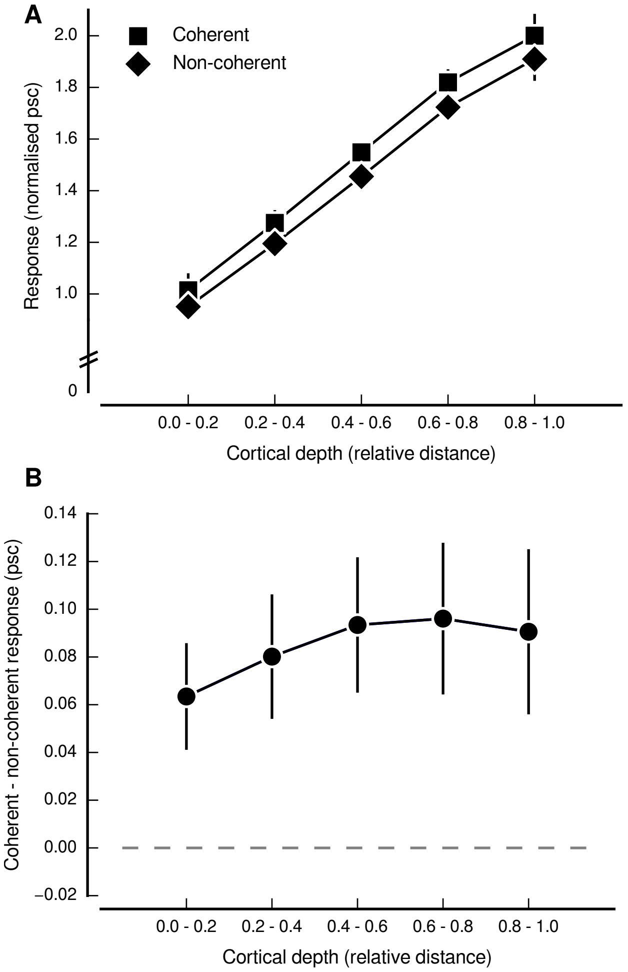


FIG. 6. Dependency of estimated response on cortical depth. **A**: Response in V1 to coherent and non-coherent image patches at different relative distances from the white matter surface. The vertical axis shows the response amplitude (percent signal change units), and the horizontal axis shows the distance from the white matter surface (20% bin width). Points show the BOLD response (normalised for differences in overall activation levels, across participants) averaged over participants, source images, and nodes on a surface at a given relative cortical depth (squares and diamonds show coherent and non-coherent conditions, respectively), and the lines are ±1 SEM. **B**: Difference between the response to coherent and non-coherent image patches at different relative distances from the white matter surface. The vertical axis shows the difference amplitude (percent signal change units), and the horizontal axis shows the distance from the white matter surface (20% bins). Points show the BOLD response difference (coherent − non-coherent) averaged over participants, source images, and nodes on a surface at a given relative cortical depth, and the lines are ±1 SEM.

The finding of a significant difference between coherent and non-coherent presentation in V1 is ostensibly in disagreement with a recent study of ours (Mannion *et al.* 2014) in which we find such differences to be confined to mid-level visual cortex. This could potentially be attributed to method differences, such as the use of block design compared to event-related design or to the analysis in a standardised versus native brain space. However, it could also be due to revealing stimulus differences between the two studies. First, we presented image patches centred at a single eccentricity in Mannion *et al.* (2014)—3°, with each patch having a diameter of 4°. As shown here, the effect of coherence seems mostly evident at more peripheral locations, raising the possibility that the stimulus layout we used previously was not within an optimal zone for eliciting contextual modulation in V1. Second, the image patches in Mannion *et al.* (2014) were presented either side of fixation along the horizontal meridian. This layout limits any contextual effects to neural circuitry that is capable of crossing between the cerebral hemispheres, whereas the contextual effects observed here in V1 could potentially be sourced from both within and between hemisphere interactions.

Despite such differences, the mid-level visual areas we identified (Mannion *et al.* 2014) could potentially be a source of the contextual modulation we observed here in V1. Purely feed-forward signaling seems unlikely to explain our results, given that each local V1 region was selected based on its independent response to a single visual field location. It is more likely that both horizontal (within V1) and feedback signals contribute to the observed contextual modulation. We speculate that the apparent peak in middle-upper regions of the cortical depth that we observed could be primarily associated with the horizontal connections that are particularly apparent in mammalian layers 2/3 (Bosking *et al.* 1997), whereas the overall difference across the cortical depths could reflect feedback influences targeting multiple cortical layers (Angelucci & Bressloff 2006).

Observers in the current study performed a concurrent behavioural task in which they were required to make a judgement about the content of the coherent scene. Are the results we report an inevitable consequence of this requirement? Signals in human V1, as measured with fMRI, are strongly affected by spatial (Gandhi *et al.* 1999) and feature-based (Saenz *et al.* 2002) attention. However, a simple application of such mechanisms appears unlikely to explain our results; an observer could not, prior to the onset of each trial, know which spatial locations or which visual features were informative regarding the coherent scene and thus neither could be targets of voluntary attention. Instead, task performance requires segmentation of the image into regions that combine for a coherent scene and those that do not. Our data indicate that V1 is a participant in this segmentation process—by representing local contextual interactions (either from horizontal or feedback sources), V1 supports the identification of spatial regions containing potential violations of typical coherent scene structure. While this process may be task-dependent, in that the circuitry may not be activated if it is unrelated to ongoing task requirements (such as when performing a demanding task at central fixation, which is likely to dampen the processing of the surrounding image structure), it is not a straightforward consequence of known attention mechanisms.

However, spatial attention may have affected the V1 responses after segmentation of the image into its coherent and non-coherent components. In this conception, contextual interactions tag each spatial location to produce separate planes for coherent and non-coherent patches—the coherent plane can then be targeted by attentional mechanisms to perform the task of classifying the coherent scene structure. The involvement of a segmentation stage is consistent with the qualitative observation that coherent and non-coherent patches are often perceived to reside at different depths, with non-coherent structure in the foreground and coherent structure in the background. Importantly, this interpretation proposes a potentially different role for V1 than that which has been espoused thus far. Instead of the increased response to aperture locations containing coherent patches reflecting pure contextual interactions, it may represent joint contributions from contextual interactions and attentional mechanisms. Given that the latter is likely to be positive and strong, the magnitude and, importantly, the sign of the contextual modulation would be highly uncertain under this interpretation.

We thus have two competing interpretations of the increased V1 response to coherent relative to non-coherent image patches that we observed. The increased response could be due to contextual interactions relating to violations of typical scene structure, or could reflect an additional contribution from attentional mechanisms operating on a segmented image. Our support for the former is rather tentative, and stems largely from the consistency of our results with those of Onat *et al.* (2013)—which were obtained with anesthetised cats and are thus unlikely to have a contribution from attention. We suggest that future studies may inform this debate by using techniques with a high temporal resolution to capitalise on the necessarily sequential involvement of contextual and attentional processes.

There are several important limitations to the current study that are also suggestive of avenues for future research. First, it is unclear whether the perceptual interpretation of a consistent scene is necessary for the effect of coherence or whether the low-level image structure is sufficient. For example, presentation of subset of the apertures would not allow the perceptual impression of coherent scene but would retain low-level interactions amongst their image structure—it is unclear if coherence would have an effect in this situation. Second, the effects we report may be limited to the particular stimulus construction that we used. Our desire to localise each aperture led us to construct a stimulus from an array of small apertures. This layout may engage cortical mechanisms that are not usually in operation during typical natural viewing. It also required us to impose a spatial scale on the contextual interactions that may not generalise to other circumstances—if each aperture was smaller or larger or the apertures were closer to or further away from each other, for example. The stimulus construction may also have determined the observed dependence on eccentricity in the difference between coherent and non-coherent responses. To stimulate approximately equally-sized regions on the cortical surface, the aperture diameter increased with eccentricity. This layout has the consequence that more peripheral apertures revealed a greater amount of image structure than apertures closer to the fovea, and the presence of this image structure may be critical to resolving the ambiguity about whether a patch belongs to the coherent scene. Finally, as we only used a limited set of images, the applicability of the findings to a novel image corpus are unknown.

In conclusion, we find that the spatial structure of the natural visual environment is an important contributor to the representation of visual information at the initial stages of cortical processing. This finding supports theories of V1 function that ascribe importance to the statistical properties of the natural visual environment, particularly those that incorporate sensitivity to the likelihood of dependency between centre and surround regions (e.g. Coen-Cagli *et al.* 2012). Furthermore, it demonstrates the multifaceted and diverse role of contextual modulation in

V1 (Nurminen & Angelucci 2014) and supports the need to further consider the functional role of contextual modulation in future studies.

## Acknowledgements

We thank A Grant, C Qui, and M-P Schallmo for scanning assistance. This work was supported by ONR (N000141210883), the WCU (World Class University) program funded by the Ministry of Education, Science, and Technology through the National Research Foundation of Korea (R31-10008), the Keck Foundation, and NIH (P30-NS076408, P41-EB015894, P30-EY011374, R21-NS075525).

## References

Allman, J., Miezin, F. & McGuinness, E. (1985) Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. *Annu Rev Neurosci*, **8**, 407–430.

Angelucci, A. & Bressloff, P. C. (2006) Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. *Prog. Brain Res.*, **154**, 93–120.

Bosking, W. H., Zhang, Y., Schofield, B. & Fitzpatrick, D. (1997) Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *J Neurosci*, **17**, 2112–2127.

Coen-Cagli, R., Dayan, P. & Schwartz, O. (2012) Cortical surround interactions and perceptual salience via natural scene statistics. *PLoS Comput Biol*, **8**, e1002405.

Cousineau, D. (2005) Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods for Psychology*, **1**, 42–45.

Cox, R. W. (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res*, **29**, 162–173.

Dale, A. M., Fischl, B. & Sereno, M. I. (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage*, **9**, 179–194.

DeAngelis, G. C., Freeman, R. D. & Ohzawa, I. (1994) Length and width tuning of neurons in the cat's primary visual cortex. *J Neurophysiol*, **71**, 347–374.

DeYoe, E. A., Carman, G. J., Bandettini, P., Glickman, S., Wieser, J., Cox, R., Miller, D. & Neitz, J. (1996) Mapping striate and extrastriate visual areas in human cerebral cortex. *Proc. Natl. Acad. Sci. USA*, **93**, 2382–2386.

Dijkstra, E. (1959) A note on two problems in connexion with graphs. *Numerische Mathematik*, **1**, 269–271.

Engel, S. A. (2012) The development and use of phase-encoded functional MRI designs. *NeuroImage*, **62**, 1195–1200.

Engel, S. A., Glover, G. H. & Wandell, B. A. (1997) Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb. Cortex*, **7**, 181–192.

Felsen, G., Touryan, J. & Dan, Y. (2005) Contextual modulation of orientation tuning contributes to efficient processing of natural stimuli. *Network*, **16**, 139–149.

Fischl, B., Sereno, M. I. & Dale, A. M. (1999) Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage*, **9**, 195–207.

Gandhi, S. P., Heeger, D. J. & Boynton, G. M. (1999) Spatial attention affects brain activity in human primary visual cortex. *Proc Natl Acad Sci U S A*, **96**, 3314–3319.

Geisler, W. S., Perry, J. S., Super, B. J. & Gallogly, D. P. (2001) Edge co-occurrence in natural images predicts contour grouping performance. *Vision Res.*, **41**, 711–724.

Gilad, A., Meirovithz, E. & Slovin, H. (2013) Population responses to contour integration: early encoding of discrete elements and late perceptual grouping. *Neuron*, **78**, 389–402.

Grinvald, A., Shoham, D., Shmuel, A., Glaser, D. E., Vanzetta, I., Shtoyerman, E., Slovin, H., Sterkin, A., Wijnbergen, C., Hildesheim, R. & Arieli, A. (1999) In-vivo optical imaging of cortical architecture and dynamics. In *Modern techniques in neuroscience research*, Springer-Verlag.

Hansen, K. A., Kay, K. N. & Gallant, J. L. (2007) Topographic organization in and near human visual area V4. *J. Neurosci.*, **27**, 11896–11911.

Larsson, J. & Heeger, D. J. (2006) Two retinotopic visual areas in human lateral occipital cortex. *J. Neurosci.*, **26**, 13128–13142.

Mannion, D. J., Kersten, D. J. & Olman, C. A. (2013) Consequences of polar form coherence for fMRI responses in human visual cortex. *Neuroimage*, **78**, 152–158.

Mannion, D. J., Kersten, D. J. & Olman, C. A. (2014) Regions of mid-level human visual cortex sensitive to the global coherence of local image patches. *J Cogn Neurosci*, **26**, 1764–1774.

Nurminen, L. & Angelucci, A. (2014) Multiple components of surround modulation in primary visual cortex: Multiple neural circuits with multiple functions? *Vision Res*, **104**, 47–56.

Olman, C. A., Harel, N., Feinberg, D. A., He, S., Zhang, P., Ugurbil, K. & Yacoub, E. (2012) Layer-specific fMRI reflects different neuronal computations at different depths in human V1. *PLoS One*, **7**, e32536.

Olmos, A. & Kingdom, F. A. A. (2004) A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception*, **33**, 1463–1473.

Onat, S., Jancke, D. & Konig, P. (2013) Cortical long-range interactions embed statistical knowledge of natural sensory input: a voltage-sensitive dye imaging study. *F1000Research*, **2**, 51.

Peirce, J. W. (2007) PsychoPy–psychophysics software in Python. *J. Neurosci. Methods*, **162**, 8–13.

Saad, Z. S., Glen, D. R., Chen, G., Beauchamp, M. S., Desai, R. & Cox, R. W. (2009) A new method for improving functional-to-structural MRI alignment using local Pearson correlation. *NeuroImage*, **44**, 839–848.

Saad, Z. S., Reynolds, R. C., Argall, B., Japee, S. & Cox, R. W. (2004) SUMA: an interface for surface-based intra- and inter-subject analysis with AFNI:. In *Proc. IEEE Int Biomedical Imaging: Nano to Macro Symp.* pages 1510–1513.

Saenz, M., Buracas, G. T. & Boynton, G. M. (2002) Global effects of feature-based attention in human visual cortex. *Nat Neurosci*, **5**, 631–632.

Schira, M. M., Wade, A. R. & Tyler, C. W. (2007) Two-dimensional mapping of the central and parafoveal visual field to human visual cortex. *J. Neurophysiol.*, **97**, 4284–4295.

Schwartz, O., Hsu, A. & Dayan, P. (2007) Space and time in visual context. *Nat Rev Neurosci*, **8**, 522–535.

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., Rosen, B. R. & Tootell, R. B. (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, **268**, 889–893.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M. & Matthews, P. M. (2004) Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, **23 Suppl 1**, S208–S219.

Vinje, W. E. & Gallant, J. L. (2000) Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, **287**, 1273–1276.

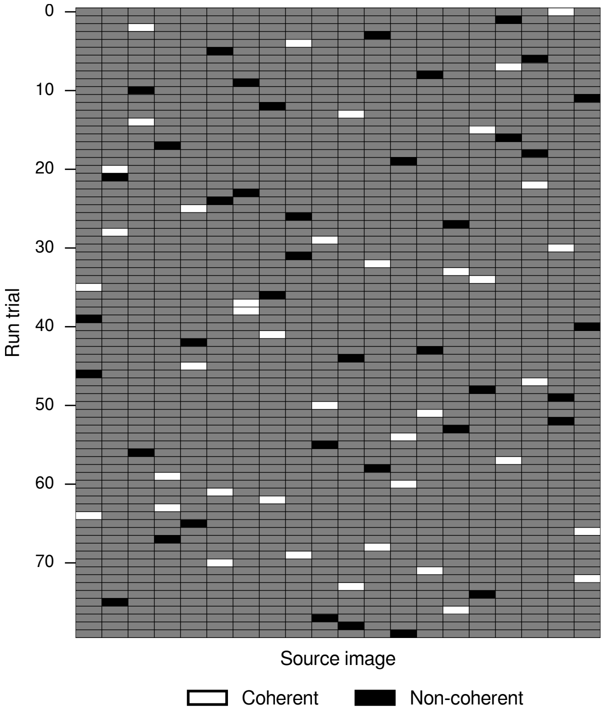
Vinje, W. E. & Gallant, J. L. (2002) Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1. *J Neurosci*, **22**, 2904–2915.

Zhao, F., Wang, P. & Kim, S.-G. (2004) Cortical depth-dependent gradient-echo and spin-echo BOLD fMRI at 9.4T. *Magn Reson Med*, **51**, 518–524.

# Supplementary Material

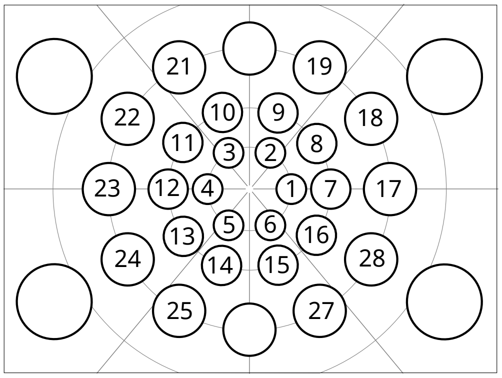
Mannion, D.J., Kersten, D.J., Olman, C.A. Scene coherence can affect the local response to natural images in human V1.

d.mannion@unsw.edu.au



**Supplementary Figure 1**. Example presentation sequence for a particular aperture for the stimulus trials in a single run. In each trial of the run, a patch is sourced from a particular image and shown within the aperture. For example, for this particular aperture and this particular run, the 19th image in the set was shown on the first trial. The source image could either be the coherent (white cells) or non-coherent (black cells) image for the trial, with the coherence depending on the source images for the other apertures. Each image was shown four times per run, equally split into coherent and non-coherent conditions.

**Supplementary Table 1.** Summary of nodes for each aperture and participant. Entries in red are those apertures or the participant that were excluded from the analysis for having insufficient nodes. The final column shows the minimum number of nodes for each patch, after excluding the participant with an overall low node count.



| | Participant | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Aperture | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Mean | Min | Min (8/9) |
| 1 | 74 | 107 | 180 | 51 | 134 | 67 | 44 | 82 | 168 | 100.78 | 44 | 44 |
| 2 | 54 | 51 | 60 | 113 | 39 | 0 | 31 | 55 | 43 | 49.56 | 0 | 31 |
| 3 | 55 | 40 | 70 | 78 | 59 | 17 | 45 | 32 | 34 | 47.78 | 17 | 32 |
| 4 | 69 | 130 | 248 | 48 | 96 | 52 | 93 | 85 | 162 | 109.22 | 48 | 48 |
| 5 | 66 | 74 | 84 | 91 | 67 | 49 | 34 | 32 | 71 | 63.11 | 32 | 32 |
| 6 | 43 | 42 | 89 | 94 | 90 | 24 | 50 | 87 | 105 | 69.33 | 24 | 42 |
| 7 | 39 | 39 | 96 | 62 | 77 | 26 | 38 | 43 | 90 | 56.67 | 26 | 38 |
| 8 | 40 | 24 | 24 | 70 | 33 | 8 | 27 | 16 | 48 | 32.22 | 8 | 16 |
| 9 | 34 | 10 | 14 | 2 | 34 | 0 | 13 | 69 | 34 | 23.33 | 0 | 2 |
| 10 | 32 | 37 | 28 | 28 | 88 | 51 | 18 | 29 | 12 | 35.89 | 12 | 12 |
| 11 | 37 | 55 | 54 | 62 | 62 | 20 | 21 | 26 | 36 | 41.44 | 20 | 21 |
| 12 | 36 | 26 | 142 | 77 | 99 | 13 | 44 | 41 | 61 | 59.89 | 13 | 26 |
| 13 | 13 | 58 | 57 | 30 | 87 | 23 | 38 | 32 | 95 | 48.11 | 13 | 13 |
| 14 | 23 | 51 | 40 | 55 | 41 | 13 | 14 | 45 | 41 | 35.89 | 13 | 14 |
| 15 | 34 | 67 | 46 | 45 | 42 | 0 | 18 | 42 | 53 | 38.56 | 0 | 18 |
| 16 | 46 | 85 | 80 | 141 | 111 | 36 | 27 | 39 | 123 | 76.44 | 27 | 27 |
| 17 | 22 | 35 | 117 | 46 | 36 | 21 | 42 | 68 | 43 | 47.78 | 21 | 22 |
| 18 | 22 | 30 | 42 | 15 | 29 | 6 | 24 | 23 | 41 | 25.78 | 6 | 15 |
| 19 | 10 | 23 | 15 | 5 | 10 | 8 | 1 | 17 | 16 | 11.67 | 1 | 1 |
| 21 | 44 | 10 | 27 | 24 | 52 | 7 | 8 | 21 | 30 | 24.78 | 7 | 8 |
| 22 | 13 | 36 | 56 | 74 | 45 | 0 | 6 | 25 | 51 | 34.00 | 0 | 6 |
| 23 | 79 | 42 | 97 | 24 | 93 | 35 | 22 | 27 | 111 | 58.89 | 22 | 22 |
| 24 | 35 | 47 | 46 | 23 | 29 | 8 | 15 | 17 | 46 | 29.56 | 8 | 15 |
| 25 | 22 | 20 | 42 | 42 | 49 | 7 | 14 | 12 | 46 | 28.22 | 7 | 12 |
| 27 | 36 | 32 | 22 | 47 | 58 | 3 | 20 | 0 | 47 | 29.44 | 0 | 0 |
| 28 | 31 | 20 | 35 | 55 | 56 | 40 | 19 | 14 | 52 | 35.78 | 14 | 14 |
| Mean | 38.81 | 45.81 | 69.65 | 53.92 | 62.15 | 20.54 | 27.92 | 37.65 | 63.81 | | | |