

Differential patterns of 2D location versus depth decoding along the visual hierarchy

Nonie J. Finlayson*, Xiaoli Zhang, and Julie D. Golomb*

Department of Psychology, Center for Cognitive & Brain Sciences, The Ohio State
University, Columbus, OH 43210, USA

*Correspondence should be addressed to Julie Golomb (golomb.9@osu.edu) and
Nonie Finlayson (nonie.j@gmail.com), Department of Psychology, The Ohio State
University, Columbus, OH 43210, USA

Abstract

Visual information is initially represented as 2D images on the retina, but our brains are able to transform this input to perceive our rich 3D environment. While many studies have explored 2D spatial representations or depth perception in isolation, it remains unknown if or how these processes interact in human visual cortex. Here we used functional MRI and multi-voxel pattern analysis to investigate the relationship between 2D location and position-in-depth information. We stimulated different 3D locations in a blocked design: each location was defined by horizontal, vertical, and depth position. Participants remained fixated at the center of the screen while passively viewing the peripheral stimuli with red/green anaglyph glasses. Our results revealed a widespread, systematic transition throughout visual cortex. As expected, 2D location information (horizontal and vertical) could be strongly decoded in early visual areas, with reduced decoding higher along the visual hierarchy, consistent with known changes in receptive field sizes. Critically, we found that the decoding of position-in-depth information tracked inversely with the 2D location pattern, with the magnitude of depth decoding gradually increasing from intermediate to higher visual and category regions. Representations of 2D location information became increasingly location-tolerant in later areas, where depth information was also tolerant to changes in 2D location. We propose that spatial representations gradually transition from 2D-dominant to balanced 3D (2D and depth) along the visual hierarchy.

Keywords: 3D space, human visual cortex, spatial representations, depth perception, fMRI multivariate pattern analysis

We live in a three dimensional (3D) world, yet visual input is initially recorded in two dimensions (2D) on the retinas. How does our visual system transform this 2D retinal input into the cohesive 3D representation of space that we effortlessly perceive? A large body of research has provided insight into how our visual systems use different cues, such as binocular disparity, perspective, shading, and motion parallax to perceive depth (Howard, 2012). What is less well understood is how position-in-depth information (hereafter referred to as depth location information) is integrated with 2D location to form a 3D perception of space.

Past research has demonstrated that 2D spatial information is represented throughout visual cortex and beyond. Both neurophysiology and functional neuroimaging studies have revealed a large number of regions in the brain sensitive to 2D visuo-spatial information: visual cortex is organized into topographic maps of 2D spatial location (Engel et al., 1994; Grill-Spector & Malach, 2004; Maunsell & Newsome, 1987; Sereno et al., 1995; Silver & Kastner, 2009; Wandell, Dumoulin, & Brewer, 2007), and 2D location information can be decoded from fMRI response patterns in early, ventral, and dorsal visual areas (Carlson, Hogendoorn, Fonteijn, & Verstraten, 2011; Fischer, Spotswood, & Whitney, 2011; Golomb & Kanwisher, 2012; Kravitz, Kriegeskorte, & Baker, 2010; Schwarzlose, Swisher, Dang, & Kanwisher, 2008).

Although often treated as a separate field, many studies have also explored how and where depth information is represented in visual cortex. Binocular disparity and/or depth-sensitive responses have been reported in several visual regions in macaques (DeAngelis & Newsome, 1999; Hubel, Wiesel, Yeagle, Lafer-Sousa, & Conway, 2015; Tsao et al., 2003) and humans (Backus, Fleet, Parker, & Heeger, 2001; Ban, Preston,

Meeson, & Welchman, 2012; Dekker et al., 2015; Durand, Peeters, Norman, Todd, & Orban, 2009; Neri, Bridge, & Heeger, 2004; Preston, Li, Kourtzi, & Welchman, 2008; Tsao et al., 2003; Welchman, Deubelius, Conrad, Bühlhoff, & Kourtzi, 2005).

Interestingly, while binocular disparity signals are found as early as V1, these signals are not thought to correspond to perception of depth until later visual areas (Barendregt, Harvey, Rokers, & Dumoulin, 2015; Cumming & Parker, 1997, 1999; Preston et al., 2008). These later visual areas (including V3A, V3B, V7, IPS, MT+, LO) have been shown to be sensitive to 3D object structure (Backus et al., 2001; Durand et al., 2009), differences in perceived depth (Neri et al., 2004; Preston et al., 2008), and the integration of different depth cues (Ban et al., 2012; Dekker et al., 2015; Murphy, Ban, & Welchman, 2013; Welchman et al., 2005). However, the nature of position-in-depth (spatial) representations remains less explored. Specifically, none of these studies have explored depth in the context of an integrated 3D representation of space, which requires combining – and comparing – information about position in depth with 2D location.

To our knowledge, our study is the first to combine and quantify both 2D and depth location information to investigate the visual representations and interactions of all three spatial dimensions. We use human functional MRI (fMRI) and multivariate pattern analysis (MVPA) to investigate how 3D spatial information is decoded throughout visual cortex. By “information”, we mean explicit, large-scale differences in neural response patterns that can be detected with fMRI MVPA. Across two experiments we explored 3D spatial representations throughout human visual cortex by comparing the amount of MVPA information about horizontal, vertical, and depth position and the

dependence/tolerance between these dimensions. The first experiment presented stimuli across the whole visual field, and was more exploratory in nature. The second experiment presented stimuli within one quadrant of the visual field, to control for possible hemifield or quadrant-based effects, and to provide a replication test for the effects found in Experiment 1.

Methods and Materials

Overview. Our approach used human fMRI to investigate how 3D spatial information is decoded in visual cortex. By 3D spatial information, we mean information about both 2D and depth location. Specifically, we refer to stimulus locations that can be defined spatially in horizontal (X), vertical (Y), and depth (Z) coordinates. We focus on the simplest case where the observer's eyes, head, and body remain stationary, and spatial position in each dimension can be expressed in terms of position relative to fixation. Observers were presented with dynamic random dot stereogram (RDS) stimuli at different 3D locations (Fig. 1A). For each participant, we used multivariate pattern analysis (Haxby et al., 2001) (MVPA) to quantify the amount of X, Y, and Z "information" that could be decoded in different parts of visual cortex. Here we measure information as explicit (measurable), large-scale differences in neural response patterns that can be detected with fMRI MVPA. It is important to note that a region's sensitivity to location information may be reliant on receptive field size and cortical magnification, such that regions with larger receptive field sizes might require larger distances between locations for position information to be decoded (Dumoulin & Wandell, 2008; Grill-Spector & Malach, 2004; Rust & DiCarlo, 2010). In the current experiment we used a fixed, equal

distance between stimuli in all three dimensions, and compared the amount of location information we were able to decode for each dimension.

The goal is to use this technique to explore broad differences in how 2D location and depth information may be organized (and interact) throughout visual cortex. We make no claims about the selectivity or preferences of individual neurons; but rather whether these preferences are organized in a spatially coherent way that would carry information detectable with a linear decoder, which is thought to reflect biologically meaningful neuronal processing (deCharms & Zador, 2000).

Participants. Each experiment included 12 participants (Experiment 1: mean age 24yrs, range 19-29yrs, 7 females; Experiment 2: mean age 23yrs, range 18-30yrs, 8 females); four participants completed both experiments. One participant was excluded from Experiment 1 due to excessive head motion. All participants had normal or corrected-to-normal vision and were screened for normal stereoacuity. Informed consent was obtained for all participants, and the Ohio State University Biomedical Sciences Institutional Review Board approved the study protocols.

Stimuli. In each experiment we stimulated 8 locations within the participants' visual field (Fig. 1A) using dynamic random dot stimuli (RDS). In Experiment 1, stimuli were small patches (2.5° square) of dynamic RDS to the left or right, above or below, and in front of or behind a central fixation point. In Experiment 2, stimulus patches were slightly smaller (1.6° square), and were located within a single visual quadrant.

The fixation point was a black circle (0.06° radius) inside a white dot (0.13° radius), superimposed on a static RDS background field (10.97° square) placed at the central

depth plane of the screen. The background field consisted of light and dark gray dots on a mid-gray background (21 dots/deg², 37% contrast). Ground and ceiling line-frames (13.4° × 3.2°) flanked this background RDS below and above, respectively, to encourage perception of a 3D space, each spanning ±20 arc min in front and behind the fixation depth plane.

The smaller dynamic RDS stimulus patches comprised black and white dots (100% contrast), with the position of the dots randomly repositioned each frame (60 Hz). For Experiment 1, the 8 stimulus locations were positioned at the corners of an invisible cube centered on the fixation point, displaced ±2.7° vertically and horizontally (3.9° eccentricity), and ±18 arc min in front or behind fixation. For Experiment 2, the stimulus locations were all positioned in the lower right quadrant of the screen, centered on a point 2.4° from fixation. The 8 locations were displaced ±1.1° vertically and horizontally from this point, and ±18 arc min in front or behind the fixation plane.

Each participant completed 8 runs of the task. Each run consisted of 19 blocks (16 stimulus blocks: 2 per location condition, and 3 fixation blocks). One location was stimulated per block for 16 s, and there was a 1.5 s inter-block gap. Location conditions were presented in a pseudo-random order, with the fixation blocks occurring at the start (block 1), middle (block 10), and end of each run (block 19). Including an extra 22.5 s of blank fixation at the end, each run lasted a total of 355 s. Participants passively viewed the stimuli while performing a dot-dimming task at fixation, detecting when the fixation frame filled into a black dot.

Depth from binocular disparity was achieved using red/green anaglyph glasses paired with Psychtoolbox's (Brainard, 1997) stereomode. The participants flipped the

glasses halfway through the experiment, after four runs, to control for low-level stimulus differences based on the color presented to each eye (the red/green color assignments were also reversed to account for this change, such that the “front” and “back” percepts were preserved, while the eye-specific color information was balanced).

Accommodation and vergence were held constant as participants maintained fixation at the same location (and depth plane) for all conditions. In a pre-screening session we confirmed that participants could accurately perceive and discriminate the two depth planes with these stimuli. Differences in perceived distance for stimuli in the 2D versus depth dimensions were measured in a supplementary psychophysics experiment (see Figure S1).

fMRI Acquisition. MRI scanning was carried out at the OSU Center for Cognitive and Behavioral Brain Imaging with a Siemens TIM Trio 3T scanner using a 32-channel receiver array head coil. Functional data were acquired with a T2-weighted gradient-echo sequence (TR = 2500 ms, TE = 28 ms, 90° flip angle). Slices were oriented to maximize coverage of the occipital, parietal, and temporal cortices (41 slices, 2×2×2 mm voxels, 10% gap). A high-resolution MPRAGE anatomical scan (1 mm³) was also acquired for each participant.

Each participant was scanned in one 2-hour session, which included the experimental runs (8 runs), functional localizers (Experiment 1: 3 runs, Experiment 2: 4 runs), and retinotopic mapping (2-6 runs each). Stimuli were generated with the Psychophysics toolbox extension (Brainard, 1997) for MATLAB (MathWorks) and displayed with a DLP projector onto a screen mounted in the rear of the scanner bore,

which participants viewed from a distance of 86 cm via a mirror at 45° above their heads attached to the head coil.

Eye Tracking. Eye position was monitored using an MR compatible Eyelink 1000 Eye Tracker, with the camera and infrared source reflected in the mirror attached to the head coil and recorded at 500 Hz. The eye tracker was calibrated at the beginning of the session and re-calibrated as necessary. Occasionally, the eye tracker signal in the scanner was too noisy to achieve reliable calibration, and the eye position was monitored via video observation.

Functional Localizers and Retinotopic Mapping. For each participant, we identified regions of interest (ROIs) using standard retinotopic mapping and functional localizer procedures. We focused on visual regions with known 2D spatial representations, as well as category-selective regions LOC and MT+ known to be sensitive to 2D location (Golomb & Kanwisher, 2012) and depth stimuli (Neri et al., 2004; Preston et al., 2008; Welchman et al., 2005). Retinotopic areas V1, V2, V3, V3A, V3B, V7, V4, and V8 were defined using rotating wedge and expanding concentric ring stimuli (Engel et al., 1994; Sereno et al., 1995). High-contrast radial checkerboard patterns were presented as 60° wedges or rings and flickered at 4 Hz. Maximal eccentricity was 16° and the central 1.6° foveal region was not stimulated (except for a central fixation point). Each run rotated clockwise or counter-clockwise or expanded or contracted through 7 cycles with a period of 24 s/cycle. Participants fixated at the center of the display and pressed a button every time the black fixation dot dimmed to gray.

Additional localizer tasks were used to identify the object-selective Lateral Occipital Complex (LOC: Kourtzi & Kanwisher, 2001; and motion-sensitive area MT+: Tootell et

al., 1995) in each participant individually. The LOC localizer task included blocks of objects and scrambled objects (Experiment 1) and objects, scrambled objects, faces, scenes, and bodies (Experiment 2) presented at the center of the screen. Participants performed a one-back repetition task, where they pressed a button whenever the exact same stimulus image was presented twice in a row. The object-selective LOC region was defined with an object > scrambled contrast. For the MT+ localizer task, participants fixated at the center of the screen and passively viewed blocks of either stationary or moving random dot displays. The stimuli were full screen dot patterns, and the moving patterns alternated between concentric motion towards and away from fixation at 7.5 Hz. The motion-sensitive MT+ area was defined with a moving > stationary contrast. We also localized an area along the intraparietal sulcus (IPS) using data from the LOC localizer task (All > Fixation contrast) in conjunction with anatomical landmarks to select a visually active region in IPS.

For some analyses, ROIs were grouped according to whether they were in dorsal or ventral streams, as well as their relative positions along the visual processing hierarchy (early visual areas V1, V2, and V3; intermediate visual areas V3A, V3B, and V4; later visual areas V7, V8, and IPS; and category selective areas MT+ and LOC). Each grouping contained both dorsal and ventral stream areas. In Experiment 2, the data were separated by hemisphere to account for the stimuli being presented only in the right visual field, and the primary analyses were conducted on data from the left hemisphere.

fMRI Preprocessing and Analysis. We used Brain Voyager QX (Brain Innovation) to preprocess the fMRI data. All data were corrected for slice acquisition time and head

motion, temporally filtered, and normalized into Talairach space (Talairach & Tournoux, 1988). Each participant's cortical surface for each hemisphere was inflated and flattened into cortical surface space for retinotopic mapping and ROI selection. Spatial smoothing of 4mm FWHM was used for the functional localizer data, but no spatial smoothing was performed on the data used for the multivariate analysis.

A whole-brain random-effects general linear model (GLM), using a canonical hemodynamic response function, was used to calculate beta weights for each voxel, for each condition and participant. For the multivariate (MVPA) analyses, separate GLMs were run for runs 1-4 ("RG"; participants had the red filter over their left eye, and green over their right) and runs 5-8 ("GR"; red/green filters flipped). Data were exported to Matlab using BrainVoyager's BVQXtools Matlab toolbox, and all subsequent analyses were done using custom code in Matlab.

Multivoxel Pattern Analysis. Multivoxel pattern analyses (MVPA) were performed for both whole-brain (searchlight) and ROI-based analyses.

ROI-based Analyses. MVPA was performed separately for each participant and ROI following the split-half method (Haxby et al., 2001), similar to Golomb and Kanwisher (Golomb & Kanwisher, 2012). To control for low-level color differences between eyes, we had participants reverse the direction of their anaglyph glasses to the opposite eyes after the first half of the experiment (between runs 4 and 5), and we conducted the split-half correlation analysis across these two halves. The data from each ROI were first split into two data sets (RG runs and GR runs). For each data set separately, the mean response across all conditions was subtracted from the responses to individual conditions, normalizing each voxel's response. Next, the voxelwise response patterns

for each of the 8 conditions in the RG run were correlated with each of the 8 conditions in the GR run, generating an 8 x 8 correlation matrix (Fig. 1B & Fig. S2). The correlations were converted to z-scores using Fisher's r-to-z transform. All subsequent analyses were performed on the z-scored data.

To quantify the amount of X, Y, and Z information contained within an ROI, along with the interactions between these dimensions, the cells in the correlation matrix were characterized according to whether they reflected the same or different position in X, Y, and Z location. For example, the Left-up-back(RG) x Left-down-back(GR) correlation would be characterized as same X, different Y, same Z (1 0 1). For each type of location, we averaged across all of the cells that reflected the "same" position, and all of the cells reflecting the "different" position (Fig. 1C), and the "same" minus "different" correlation difference was taken as a measure of the amount of "information" about that property. E.g., X-information was quantified as the difference in correlation between all conditions that shared the same X position (1 - -) versus differed in X position (0 - -). This standard approach (Haxby et al., 2001) is based on the rationale that if an ROI contains information about a certain type of location, then the voxelwise response pattern should be more similar for two conditions that share the same location than differ in location. This approach is an alternative to calculating decoding accuracy using machine-learning techniques, and generally produces highly similar patterns of results to those obtained from support vector machines (SVM) (Golomb & Kanwisher, 2012).

Searchlight Analyses. MVPA "searchlight" analyses (Kriegeskorte, Goebel, & Bandettini, 2006) were performed using the approach described above to search across our entire slice coverage for voxel clusters showing significant X, Y, and Z information.

Our slice prescription included full coverage of occipital and parietal cortices and posterior coverage of temporal and frontal cortices, but did not cover regions such as prefrontal cortex. For each participant, we iteratively searched through the brain conducting MVPA within a “moving” ROI defined as a sphere of radius 3 mm (~100 voxels). On each iteration, the ROI was chosen as a sphere centered on a new voxel, and multivoxel correlation analyses were performed exactly as described above. The magnitudes of X, Y, and Z information (as defined by the z-transformed “same” – “different” correlation differences) were then plotted for each voxel, creating a z-map for each type of information for each participant. These maps were then spatially smoothed with a 3 mm FWHM kernel and combined across subjects using one-sample t-tests to identify clusters containing significant information about each property. The resulting t-maps were thresholded at $p < .05$, and familywise error rate was controlled with BrainVoyager cluster correction at 25 voxels.

Hybrid Searchlight. We also conducted an additional searchlight analysis to directly compare 2D versus depth location decoding across the brain. To do this, we first created a 2D (XY) information map for each participant by averaging the amount of X and Y information (z-transformed correlation differences). The XY and Z maps were then averaged across participants and compared by subtracting the magnitude of Z information from the magnitude of XY information for each voxel. The resulting difference maps were thresholded to include only voxels that showed significant ($p < .05$, cluster corrected) location information for at least one of the dimensions (XY and/or Z). This thresholding criteria was chosen to ensure that voxels exhibiting near-zero

values in the difference map were voxels in which *both* XY and Z information were present (balanced 3D representation) as opposed to *neither* (no location information).

Tolerance Analyses. To assess whether the X, Y, and Z information in a given ROI was tolerant of or dependent on the other dimensions, correlation differences were calculated on different subsets of the data. E.g., using the same [X Y Z] coding system as above, we would calculate the amount of “dependent” X information as the difference in same-X minus different-X when both Y and Z were the same ($[1\ 1\ 1] - [0\ 1\ 1]$), and the amount of “tolerant” X-information as the difference in same-X minus different-X when both Y and Z were different ($[1\ 0\ 0] - [0\ 0\ 0]$). We then calculated a location-tolerance index for each dimension (Fig. 4 inset):

$$\text{Tolerance Index} = 1 - \frac{z(\text{dependent}) - z(\text{tolerant})}{z(\text{dependent}) + z(\text{tolerant})}$$

A larger index (close to 1) means that the selected dimension is highly tolerant of changes in location along the other dimensions. A smaller index (close to 0) indicates the location information is highly dependent on the other dimensions. Because individual subjects had occasional negative values for the tolerant or dependent scores, it would have been problematic to calculate this index for each subject (Simmons, Bellgowan, & Martin, 2007), so we calculated the tolerance index on the group-averaged data only.

Representational Similarity Analysis (RSA) and Multidimensional Scaling (MDS). To analyze the similarity between the full pattern of location information across brain regions, we conducted a Representational Similarity Analysis (RSA; Kriegeskorte et al.

2008). First, we created 8 x 8 correlation matrices for each participant for each of our 11 ROIs based on the full set of data (not split-halves). We z-transformed the correlation data and averaged across all participants to get a single matrix for each ROI. We then correlated these correlation matrices with each other to create an 11 x 11 Representational Similarity Matrix across ROIs. Each cell was calculated as the distance (or dissimilarity) between a pair of ROIs (quantified by $1 - r$). We then used Multidimensional Scaling (MDS; Kruskal & Wish, 1978) to calculate a set of inter-point distances in N-dimensional space and visualize how ROIs clustered together in similarity space.

Results and Discussion

Whole-brain comparison of X, Y, and Z location information

We first conducted an MVPA “searchlight” analysis (Kriegeskorte et al., 2006) to test where in the brain horizontal (X), vertical (Y), and depth (Z) information could be decoded (Fig. 2A). Searchlight maps were largely consistent across Experiments 1 and 2, with the exception that X information was more widespread in Experiment 1, likely reflecting regions that exhibit broad contralateral, hemisphere-based information.

As expected, most of the location information for all three dimensions was found in visual cortex. In both experiments there was widespread coverage of X and Y location information throughout visual cortex, whereas Z information was absent in the earliest (most posterior) visual areas and only appeared in intermediate and later regions. The Z information was spread throughout both dorsal and ventral visual areas, and largely

overlapped with the X and Y coverage. Additionally, Z information could be decoded in some higher parietal regions, as well as a region in the frontal cortex possibly corresponding to Frontal Eye Fields (FEF; Ferraina, Paré, & Wurtz, 2000). Although there were some parietal and frontal clusters for X and Y information, they tended to be less consistent across experiments, and this area was outside of the slice prescription for some participants.

Evidence for a gradual transition

The searchlight results described above suggest that the relative amounts of 2D and depth information may vary along the visual hierarchy, with only 2D information decodable in the earliest visual areas but all three dimensions decodable in later visual areas. Might visual cortex contain a gradual transition from 2D-dominant to balanced 3D (2D and depth) spatial representations? To more directly explore this question, we next conducted a hybrid searchlight analysis comparing the relative amounts of X/Y versus Z information. We focused this analysis on the within-quadrant Experiment 2 design. We first created a single 2D map (XY average) and then subtracted the Z information map to visualize relative differences in the amount of 2D versus depth information that could be decoded (Fig. 2B). The results revealed a striking gradient. The most posterior occipital regions were very heavily weighted toward XY information (2D), but this weighting gradually shifted moving into intermediate and later visual areas, which exhibited increasingly balanced decoding of all three dimensions (roughly equal XYZ information). Finally, certain higher-level areas were weighted more heavily toward the

depth dimension (more Z information than XY information), though these clusters appear more isolated.

ROI analysis: Comparison of 2D vs depth location decoding across visual cortex

The searchlight results suggest a gradual, systematic transition from 2D-dominant to balanced 3D (2D and depth) spatial information along the visual hierarchy – but is this transition driven simply by reduced decoding of 2D information? To quantify and test this account, we identified multiple ROIs for each participant using standard retinotopic mapping and functional localizer procedures (see figure S3 for ROI locations relative to the searchlight), and examined the amounts of X, Y, and Z information that could be decoded in each of these ROIs. Results for each of the individual ROIs are shown in Figure S4 and Table S1. All regions demonstrated significant X and Y location information decoding, and significant Z decoding was found in both experiments in regions V3A, V7, IPS, and MT+ (for full breakdown see Tables 1-3).

Visual hierarchy effects. To test if the gradual pattern seen in the searchlight was driven by a decrease in 2D information along the hierarchy, an increase in depth information, or both, we classified the ROIs into four groups (V1/V2/V3, V3A/V3B/V4, V7/V8/IPS, MT+/LOC), according to their relative location in the standard visual hierarchy (Felleman & van Essen, 1991), with each group containing both dorsal and ventral stream areas.

Figures 3A and 3B illustrate how the patterns of X, Y, and Z information changed across the visual hierarchy. Several patterns become apparent. First, in both experiments, the earliest visual areas contained almost exclusively 2D information, in

line with our searchlight results. Second, in Experiment 1, the amount of X information remained high across all ROI groups, dominating the response even in later areas. As discussed earlier, this effect is likely driven by large contralateral, hemisphere-based preferences that persist throughout visual cortex (Carlson et al., 2011; Hemond, Kanwisher, & Op de Beeck, 2007); indeed, this effect was mitigated in Experiment 2, when the stimulus locations were all presented within a single quadrant. Third, and most notably, the amount of Z information gradually increased along the hierarchy in both experiments.

It should be emphasized that the critical question here is not whether there is overall more X and Y information than Z information, but whether the relationship between these dimensions changes along the hierarchy. Two-way repeated-measures ANOVAs with Hierarchy (V1/V2/V3, V3A/V3B/V4, V7/V8/IPS, MT+/LOC) and Dimension (X, Y, Z) revealed a significant interaction between hierarchy and spatial dimension in both experiments (Experiment 1, $F_{6,60} = 27.16$, $p < .001$, $\eta_p^2 = .73$; Experiment 2, $F_{6,66} = 11.25$, $p < .001$, $\eta_p^2 = .51$). Follow up tests revealed that this interaction was driven by *both* a decrease in 2D information (Experiment 1: $F_{3,30} = 65.65$, $p < .001$, $\eta_p^2 = .87$; Experiment 2: X: $F_{3,33} = 17.48$, $p < .001$, $\eta_p^2 = .61$; Y: $F_{3,33} = 11.45$, $p < .001$, $\eta_p^2 = .51$), and an increase in Z information ($F_{3,33} = 5.89$, $p = .002$, $\eta_p^2 = .35$) along the hierarchy. Importantly, these results demonstrate that the transition is not driven solely by a decrease in 2D information. This overall pattern of results was largely similar in both dorsal and ventral streams (Supplemental Analysis S1).

Interactions between spatial dimensions

How dependent/tolerant is each dimension to changes in the other dimensions? Our results so far have focused on overall location information for each of the three dimensions. However, our design also allows us to explore important questions about the interactions between dimensions. For example, can depth information be decoded even when X and Y location are different?

Figure 4 shows X, Y, and Z information broken down in terms of how tolerant or dependent it was on the other two dimensions for Experiment 2 (see Fig. S7 for additional interaction analyses). First, as noted earlier, the overall amount of X and Y location information decreased along the hierarchy while Z information increased (Dimension x Hierarchy interaction: $F_{6,66} = 11.25$, $p < .001$, $\eta_p^2 = .51$). Interestingly, the tolerance varied as well (Tolerance x Hierarchy interaction: $F_{3,33} = 9.78$, $p < .001$, $\eta_p^2 = .47$; 3-way interaction: $F_{6,66} = 12.82$, $p < .001$, $\eta_p^2 = .54$). Follow-up tests found this to be primarily driven by an increase in tolerance for X and Y (Tolerance x Hierarchy interaction for X: $F_{3,33} = 7.44$, $p < .001$, $\eta_p^2 = .57$; Y: $F_{3,33} = 21.9$, $p < .001$, $\eta_p^2 = .67$; Z: $F_{3,33} = 0.89$, $p = .457$, $\eta_p^2 = .08$). As illustrated by Figure 4 inset, X and Y location information were moderately dependent on the other dimensions in early visual cortex, and became relatively more tolerant in higher visual areas, consistent with prior reports (Carlson et al., 2011; Rust & DiCarlo, 2010). Interestingly, when Z location information was present, it was overall more tolerant to location changes in the other dimensions than X or Y information was.

Are these spatial representations globally decoded? A related question is whether these representations are global – that is, can information about the horizontal, vertical, or depth position of a stimulus be decoded in the un-stimulated hemisphere? Figure S5

illustrates decoding for each dimension in Experiment 2 in the contralateral and ipsilateral hemispheres. Overall the amount of location information was substantially weaker in the ipsilateral hemisphere. In some areas, X and Y information could still be decoded above chance, though receptive fields that occasionally extend across the meridian might drive this. Strikingly, no Z information could be decoded in any region from the ipsilateral hemisphere (Table S3). Thus, while depth information seems to be tolerant to changes in X and Y location, it is not global.

Representational similarity across visual cortex. As a final exploratory analysis, we conducted a Representational Similarity Analysis (RSA; Kriegeskorte et al. 2008) with Multidimensional Scaling (MDS; Kruskal & Wish, 1978), comparing the full 8x8 pattern of data for all of our regions of interest. Most regions generally clustered with other regions similarly positioned in the visual hierarchy (Fig. 5), consistent with the hypothesis that the relationship between 2D and depth location information varies along the visual hierarchy.

General Discussion

Our study provides the first direct investigation of the interactions between 2D location and position-in-depth information in human visual cortex. While many previous studies have explored the decoding of 2D location information in different visual areas, here our focus was on how the decoding of depth location varies along the visual hierarchy, particularly with respect to how it compares to (and interacts with) 2D location information. We found that depth location information was not reliably decoded in the earliest visual regions, but gradually increased into intermediate and higher-level

regions, while 2D location information simultaneously decreased. Rather than a few isolated position-in-depth “modules” or decoding differentiating along dorsal/ventral visual streams, our results are most consistent with a widespread, gradual transition along the visual hierarchy.

2D location information

Most studies that have explored location representations in the brain have focused on 2D location (Carlson et al., 2011; Fischer et al., 2011; Golomb & Kanwisher, 2012; Kravitz et al., 2010; Schwarzlose et al., 2008), and our general findings regarding 2D location information are consistent with this prior literature. We find that 2D location information is present in all visual areas and decreases in magnitude (or sensitivity) along the visual hierarchy (except when horizontal locations are divided across hemisphere). These findings fit with evidence that receptive fields become larger and less precise along the hierarchy (Dumoulin & Wandell, 2008; Grill-Spector & Malach, 2004; Rust & DiCarlo, 2010), although contralateral bias may remain (Carlson et al., 2011; Hemond et al., 2007).

Depth representations

A number of studies have looked at how different aspects of depth are represented in human visual cortex (Backus et al., 2001; Ban et al., 2012; Dekker et al., 2015; Neri et al., 2004; Preston et al., 2008; Tsao et al., 2003; Welchman et al., 2005). Most have focused on non-spatial aspects of depth, e.g., neural representations of 3D object structure (Backus et al., 2001; Durand et al., 2009), or the integration of different depth

cues (Ban et al., 2012; Dekker et al., 2015; Murphy et al., 2013; Welchman et al., 2005), although a few recent studies have examined fMRI sensitivity to differences in depth from disparity, finding regions that are sensitive to absolute vs relative (Neri et al., 2004) or metric vs categorical (Preston et al., 2008) depth differences. Neurophysiology studies have also reported neurons with different depth preferences in various visual areas (DeAngelis & Newsome, 1999; Hubel et al., 2015; Tsao et al., 2003). We found significant position-in-depth information in similar regions, including V3A, V7, IPS, and MT+. Yet strikingly, no research had addressed the fundamental question of if/how depth information compares and interacts with 2D spatial information. Whereas most depth studies had used large-field stimuli consisting of front versus back depth planes spanning most of the visual field (e.g., Ban et al, 2012), our stimuli varied in all three dimensions, allowing us to measure 2D and depth information for the same stimuli, as well as comparing sensitivities to depth information across 2D locations. Furthermore, while most studies have focused on whether there is any significant depth information in a given region, our approach allows us to explore possible large-scale organization schemes across visual cortex. Indeed, the finding that explicit depth information gradually increases from early to later visual cortex while 2D information decreases suggests a gradient of depth information throughout visual cortex. We also find that this depth information is relatively tolerant to changes in 2D location, indicating that those regions may represent whether an object is near or far regardless of where in 2D space it is.

It is worth noting that while neurons in early visual areas respond to binocular disparity, here the test was whether the fMRI pattern could differentiate between relative

disparities of equal size but opposite sign, to explore representation of position-in-depth. It is possible that our results reflect neural representations of disparity sign, rather than perceived depth per se, but this seems unlikely given our pattern of results and the lack of depth decoding in early visual cortex. Neurons in early visual cortex have been shown to respond to binocular disparity without depth perception (Cumming & Parker, 1997, 1999), but perceptually-relevant depth representations tend to not emerge until later. For example, one recent fMRI study found that integrated binocular perception (cyclopean representation) does not emerge until after V1 (Barendregt et al., 2015), and another study using MVPA also found that representations in early visual cortex were not directly related to perceptual estimates of depth, whereas several intermediate and later visual areas exhibited preferential decoding for perceptually-relevant depth from disparity cues (correlated versus anti-correlated stimuli; (Preston et al., 2008). Furthermore, because the binocular disparities we used were small horizontal location shifts in opposite directions in each eye, we might predict that if our results were driven by decoding these horizontal differences, the pattern of decoding should mimic what we found for the horizontal dimension, decreasing in magnitude along the hierarchy in Experiment 2, but instead we found the opposite pattern. Although we are not able to definitively generalize our results to the representation of depth from other cues, our results lay a crucial foundation for understanding how the brain might integrate information about an object's position in depth with 2D location on the retina to form perceptually relevant 3D spatial representations.

Transition from 2D to 3D along hierarchy

The most notable conclusion from our findings is how the relationship between 2D location and position-in-depth information changes along the visual hierarchy. Although the three dimensions of location information are similar in some ways – e.g., all become location-tolerant in later visual areas – the pattern of decoding along the visual hierarchy varies strikingly. While decoding of 2D location information decreases, depth location decoding increases along the visual hierarchy. This supports the interesting possibility that spatial representations in visual cortex gradually transition from 2D-dominant to balanced 3D (2D and depth). Interestingly, while our results reveal that the three dimensions may be equally decodable in these later areas, the tolerance data suggests that the three dimensions are at least somewhat independent of each other in these later visual regions.

Such a transition makes sense given that the visual system is organized hierarchically and might be expected to shift from more simple to more complex visual processing (Felleman & van Essen, 1991; Grill-Spector & Malach, 2004). However, widespread transitions are not found for all aspects of visual processing. For example, similar transitions have been hypothesized for other types of visuo-spatial information, notably the question of whether visual cortex transitions from native retinotopic (eye-centered) spatial representations to more perceptually relevant spatiotopic (world-centered) representations. However, a previous paper (Golomb & Kanwisher, 2012) using the same approach as here failed to find any evidence of a transition from early to later areas, instead finding that even higher-level visual areas still contain a purely retinotopic representation. The current results are somewhat surprising in this context, and raise the interesting suggestion that depth information may be more fundamental

than spatiotopic information. While the visual system appears to adopt a strategy of continually updating spatiotopic position on the fly (rather than converting 2D retinotopic information into explicit spatiotopic representations), the visual system does seem to transform 2D information into explicit representations of depth position in later visual areas. One reason this might be is that position-in-depth may be more likely to be coded relative to the self, in an egocentric reference frame. Because our 3D stimuli were eye- and head-centric, the position-in-depth information may reflect differences relative to the fixation plane, which in a sense may be more analogous to “retinotopic” representations. Further research would be needed to investigate whether the 2D to 3D transition we report here holds for the representation of absolute position-in-depth, following up on other studies exploring relative versus absolute disparity (Cumming & Parker, 1997, 1999; Neri et al., 2004).

Why didn't we find depth information in early visual areas, given that that disparity information is present in early visual cortex (Ban et al., 2012; Dekker et al., 2015; Poggio, Gonzalez, & Krause, 1988; Preston et al., 2008)? First, it is important to note the difference between disparity information and position-in-depth. As discussed above, while binocular disparity signals are found as early as V1, these signals are not thought to correspond to perception of depth until later visual areas (Barendregt et al., 2015; Cumming & Parker, 1997, 1999; Preston et al., 2008). In our study, we compared one direction of disparity with an equal but opposite direction of disparity, hence finding no representation of binocular disparity itself. Of course, it is still possible that position-in-depth information also exists in earlier visual regions, just at a finer or more spatially distributed scale than can be detected with these techniques (Freeman, Brouwer,

Heeger, & Merriam, 2011; Op de Beeck, 2010). It is also worth noting that the amount of “information” we can decode with MVPA may be dependent on the presented stimulus distances. The increase in depth information we report could be driven by an increase in the number of depth-sensitive neurons, an increase in selectivity or sensitivity of individual neurons, and/or an increase in the spatial separation of neurons with different depth preferences (resulting in a more detectable population response). The same could be said for the decrease in 2D location information, although it is worth noting that we see the same pattern of results in Experiments 1 and 2, which had different stimulus sizes. In other words, it’s possible that individual neurons in a given area might respond just as strongly to depth information and 2D information, but this location information may be organized differently, resulting in different patterns of decoding. Crucially, it is clear that (1) the representation of depth information is changing along the visual hierarchy, becoming increasingly detectable (explicitly decodable) in the large-scale pattern of fMRI response, and (2) this pattern is in direct contrast to the reduced decoding seen along the hierarchy for 2D spatial information.

Our results suggest that spatial representations shift from primarily 2D to balanced 3D along the hierarchy, although there may be some alternative explanations for this transition. One possibility is that attentional effects may drive the decoding of Z information. E.g., attending more to front than back stimuli (or vice versa) could result in overall signal differences that might inflate decoding, particularly as attentional effects are known to increase along the hierarchy (Roberts, Allen, Dent, & Humphreys, 2015). However, we conducted a univariate (mean response magnitude) analysis in addition to our MVPA analysis (see Fig. S8), and found a mix of both front-preferring and back-

preferring regions, arguing against an overall attentional bias. Another possibility is that the depth representations may not necessarily reflect spatial information in the same way as 2D spatial information, but rather that depth is being represented more as a feature of an object. In the past, the investigation of depth has often focused on 3D object structure (Todd, 2004; Welchman et al., 2005), though behavioral studies have demonstrated 3D position to be important for perception (Aks & Enns, 1996; Finlayson & Grove, 2015). The current results cannot conclusively answer whether depth is a spatial dimension or a feature (or whether this differs across brain regions), but they provide a crucial first step in characterizing the nature of depth position information relative to 2D information, and how these signals might interact to form a 3D representation of space.

Acknowledgements: This work was supported by research grants from the National Institutes of Health (R01-EY025648 to J.G.) and Alfred P. Sloan Foundation (BR-2014-098 to J.G.). We thank C. Kupitz and A. Shafer-Skelton for assistance with programming and subject testing; J. Mattingley, J. Todd, B. Harvey, J. Fischer, and A. Leber for helpful discussion; and the OSU Center for Cognitive and Behavioral Brain Imaging for research support.

References

- Aks, D. J., & Enns, J. T. (1996). Visual search for size is influenced by a background texture gradient. *Journal of Experimental Psychology: Human Perception and Performance*, 22(6), 1467.
- Backus, B. T., Fleet, D. J., Parker, A. J., & Heeger, D. J. (2001). Human Cortical Activity Correlates With Stereoscopic Depth Perception. *Journal of Neurophysiology*, 86(4), 2054–2068.
- Ban, H., Preston, T. J., Meeson, A., & Welchman, A. E. (2012). The integration of motion and disparity cues to depth in dorsal visual cortex. *Nature Neuroscience*, 15(4), 636–643. <https://doi.org/10.1038/nn.3046>
- Barendregt, M., Harvey, B. M., Rokers, B., & Dumoulin, S. O. (2015). Transformation from a Retinal to a Cyclopean Representation in Human Visual Cortex. *Current Biology*. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0960982215006685>
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436.

- Carlson, T., Hogendoorn, H., Fonteijn, H., & Verstraten, F. A. J. (2011). Spatial coding and invariance in object-selective cortex. *Cortex*, *47*(1), 14–22.
<https://doi.org/10.1016/j.cortex.2009.08.015>
- Cumming, B. G., & Parker, A. J. (1997). Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature*, *389*(6648), 280–283.
<https://doi.org/10.1038/38487>
- Cumming, B. G., & Parker, A. J. (1999). Binocular neurons in V1 of awake monkeys are selective for absolute, not relative, disparity. *The Journal of Neuroscience*, *19*(13), 5602–5618.
- DeAngelis, G. C., & Newsome, W. T. (1999). Organization of Disparity-Selective Neurons in Macaque Area MT. *The Journal of Neuroscience*, *19*(4), 1398–1415.
- deCharms, R. C., & Zador, A. (2000). Neural Representation and the Cortical Code. *Annual Review of Neuroscience*, *23*(1), 613–647.
<https://doi.org/10.1146/annurev.neuro.23.1.613>
- Dekker, T. M., Ban, H., van der Velde, B., Sereno, M. I., Welchman, A. E., & Nardini, M. (2015). Late Development of Cue Integration Is Linked to Sensory Fusion in Cortex. *Current Biology*, *25*(21), 2856–2861.
<https://doi.org/10.1016/j.cub.2015.09.043>
- Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *NeuroImage*, *39*(2), 647–660.
<https://doi.org/10.1016/j.neuroimage.2007.09.034>

- Durand, J.-B., Peeters, R., Norman, J. F., Todd, J. T., & Orban, G. A. (2009). Parietal regions processing visual 3D shape extracted from disparity. *NeuroImage*, *46*(4), 1114–1126. <https://doi.org/10.1016/j.neuroimage.2009.03.023>
- Engel, S. A., Rumelhart, D. E., Wandell, B. A., Lee, A. T., Glover, G. H., Chichilnisky, E.-J., & Shadlen, M. N. (1994). fMRI of human visual cortex. *Nature*. Retrieved from <http://psycnet.apa.org/psycinfo/1995-00647-001>
- Felleman, D. J., & van Essen, D. C. (1991). Distributed Hierarchical Processing in the Primate. *Cerebral Cortex*, *1*(1), 1–47. <https://doi.org/10.1093/cercor/1.1.1>
- Ferraina, S., Paré, M., & Wurtz, R. H. (2000). Disparity sensitivity of frontal eye field neurons. *Journal of Neurophysiology*, *83*(1), 625–629.
- Finlayson, N. J., & Grove, P. M. (2015). Visual search is influenced by 3D spatial layout. *Attention, Perception, & Psychophysics*, 1–9. <https://doi.org/10.3758/s13414-015-0924-3>
- Fischer, J., Spotswood, N., & Whitney, D. (2011). The emergence of perceived position in the visual system. *Journal of Cognitive Neuroscience*, *23*(1), 119–136.
- Freeman, J., Brouwer, G. J., Heeger, D. J., & Merriam, E. P. (2011). Orientation Decoding Depends on Maps, Not Columns. *The Journal of Neuroscience*, *31*(13), 4792–4804. <https://doi.org/10.1523/JNEUROSCI.5160-10.2011>
- Golomb, J. D., & Kanwisher, N. (2012). Higher level visual cortex represents retinotopic, not spatiotopic, object location. *Cerebral Cortex*, *22*(12), 2794–2810.
- Grill-Spector, K., & Malach, R. (2004). The Human Visual Cortex. *Annual Review of Neuroscience*, *27*(1), 649–677. <https://doi.org/10.1146/annurev.neuro.27.070203.144220>

- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science*, *293*(5539), 2425–2430.
<https://doi.org/10.1126/science.1063736>
- Hemond, C. C., Kanwisher, N. G., & Op de Beeck, H. P. (2007). A Preference for Contralateral Stimuli in Human Object- and Face-Selective Cortex. *PLoS ONE*, *2*(6), e574. <https://doi.org/10.1371/journal.pone.0000574>
- Howard, I. P. (2012). *Perceiving in Depth Volume 1 Basic Mechanisms*. Oxford University Press. Retrieved from
<http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199764143.001.0001/acprof-9780199764143>
- Hubel, D. H., Wiesel, T. N., Yeagle, E. M., Lafer-Sousa, R., & Conway, B. R. (2015). Binocular stereoscopy in visual areas V-2, V-3, and V-3A of the macaque monkey. *Cerebral Cortex*, *25*(4), 959–971.
- Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science*, *293*(5534), 1506–1509.
- Kravitz, D. J., Kriegeskorte, N., & Baker, C. I. (2010). High-Level Visual Object Representations Are Constrained by Position. *Cerebral Cortex*, *20*(12), 2916–2925. <https://doi.org/10.1093/cercor/bhq042>
- Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proc Natl Acad Sci U S A*, *103*(10), 3863–8.
<https://doi.org/10.1073/pnas.0600244103>

- Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis—connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2605405/>
- Kruskal, J. B., & Wish, M. (1978). *Multidimensional Scaling*. SAGE.
- Maunsell, J. H., & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience*, 10(1), 363–401.
- Murphy, A. P., Ban, H., & Welchman, A. E. (2013). Integration of texture and disparity cues to surface slant in dorsal visual cortex. *Journal of Neurophysiology*, 110(1), 190–203.
- Neri, P., Bridge, H., & Heeger, D. J. (2004). Stereoscopic Processing of Absolute and Relative Disparity in Human Visual Cortex. *Journal of Neurophysiology*, 92(3), 1880–1891. <https://doi.org/10.1152/jn.01042.2003>
- Op de Beeck, H. P. (2010). Probing the mysterious underpinnings of multi-voxel fMRI analyses. *NeuroImage*, 50(2), 567–571.
- Poggio, G. F., Gonzalez, F., & Krause, F. (1988). Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity. *The Journal of Neuroscience*, 8(12), 4531–4550.
- Preston, T. J., Li, S., Kourtzi, Z., & Welchman, A. E. (2008). Multivoxel Pattern Selectivity for Perceptually Relevant Binocular Disparities in the Human Brain. *The Journal of Neuroscience*, 28(44), 11315–11327. <https://doi.org/10.1523/JNEUROSCI.2728-08.2008>

- Roberts, K. L., Allen, H. A., Dent, K., & Humphreys, G. W. (2015). Visual search in depth: The neural correlates of segmenting a display into relevant and irrelevant three-dimensional regions. *NeuroImage*, *122*, 298–305.
- Rust, N. C., & DiCarlo, J. J. (2010). Selectivity and tolerance (“invariance”) both increase as visual information propagates from cortical area V4 to IT. *The Journal of Neuroscience*, *30*(39), 12978–12995.
- Schwarzlose, R. F., Swisher, J. D., Dang, S., & Kanwisher, N. (2008). The distribution of category and location information across object-selective regions in human visual cortex. *Proceedings of the National Academy of Sciences*, *105*(11), 4447–4452. <https://doi.org/10.1073/pnas.0800431105>
- Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., ... Tootell, R. B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, *268*(5212), 889–93.
- Silver, M. A., & Kastner, S. (2009). Topographic maps in human frontal and parietal cortex. *Trends in Cognitive Sciences*, *13*(11), 488–495.
- Simmons, W. K., Bellgowan, P. S., & Martin, A. (2007). Measuring selectivity in fMRI data. *Nature Neuroscience*, *10*(1), 4–5.
- Talairach, J., & Tournoux, P. (1988). Co-planar stereotaxic atlas of the human brain. 3-Dimensional proportional system: an approach to cerebral imaging. Retrieved from <http://www.citeulike.org/group/96/article/745727>
- Todd, J. T. (2004). The visual perception of 3D shape. *Trends in Cognitive Sciences*, *8*(3), 115–121.

Tootell, R. B., Reppas, J. B., Kwong, K. K., Malach, R., Born, R. T., Brady, T. J., ...

Belliveau, J. W. (1995). Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *Journal of Neuroscience*, 15(4), 3215.

Tsao, D. Y., Vanduffel, W., Sasaki, Y., Fize, D., Knutsen, T. A., Mandeville, J. B., ...

Van Essen, D. C. (2003). Stereopsis activates V3A and caudal intraparietal areas in macaques and humans. *Neuron*, 39(3), 555–568.

Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007). Visual field maps in human cortex. *Neuron*, 56(2), 366–383.

Welchman, A. E., Deubelius, A., Conrad, V., Bühlhoff, H. H., & Kourtzi, Z. (2005). 3D

shape perception from combined depth cues in human visual cortex. *Nature Neuroscience*, 8(6), 820–827. <https://doi.org/10.1038/nn1461>

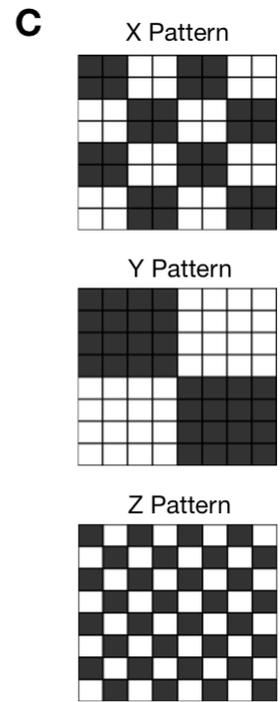
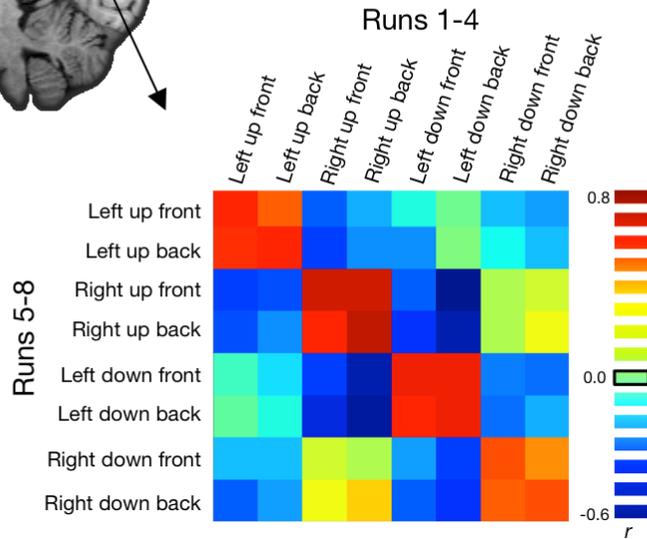
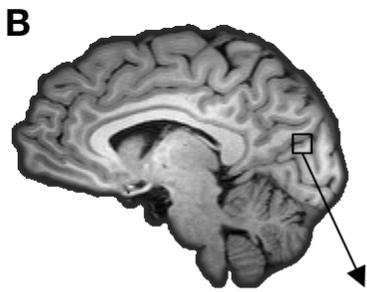
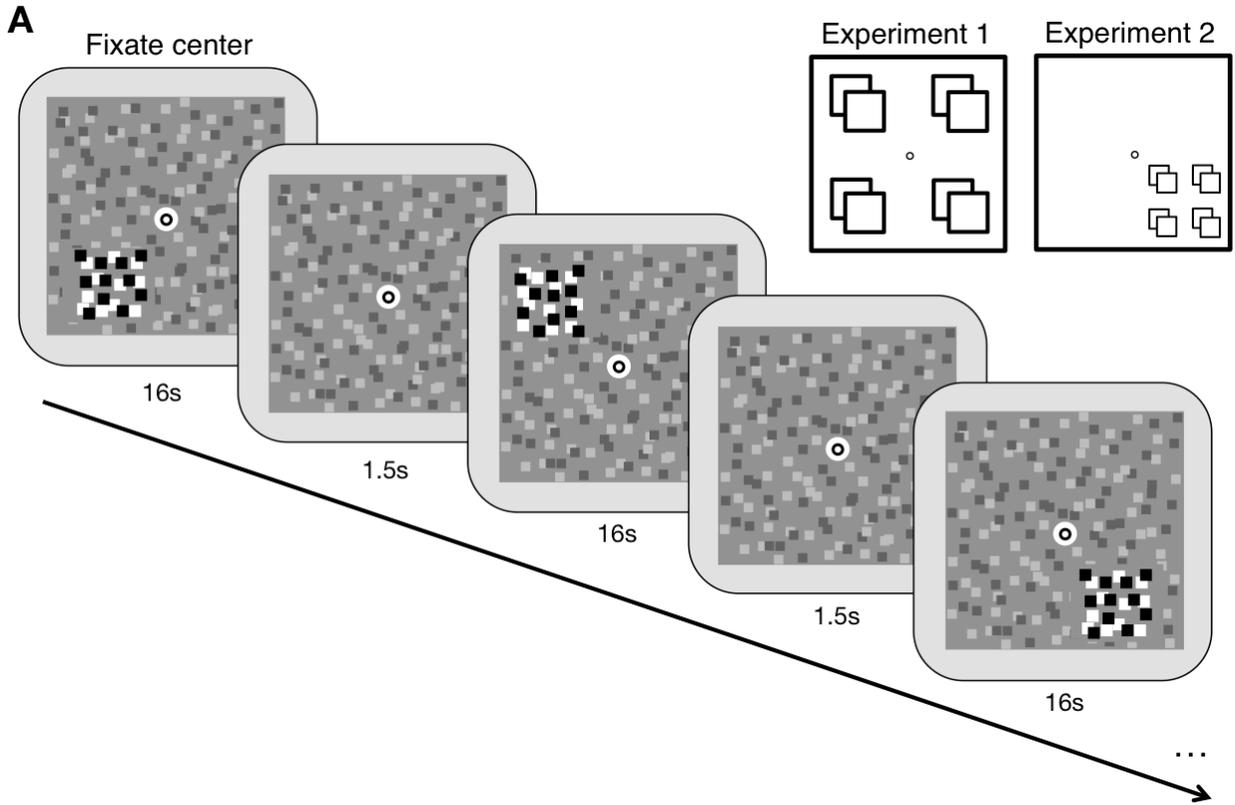


Figure 1. A, Schematic illustration of the stimuli. Stimuli were high contrast dynamic random dot patterns on a lower contrast static random dot background. Stimuli were presented in a block design, with two blocks per each of the eight location conditions per run. Each block lasted 16 s with 1.5 s between each block. Participants performed a dot-dimming task at fixation, pressing a button whenever the fixation dot flashed black. Inset illustrates the possible stimulus locations. For Experiment 1 locations were either to the left or right of fixation, above or below fixation, and in front of or behind fixation. In Experiment 2 all stimulus locations were in the lower right quadrant of the display; X and Y distances were smaller, but Z distance was the same as Experiment 1. **B**, Correlation matrices were created for each searchlight or ROI (shown here: actual data from Experiment 1 V7). This matrix is created by correlating the voxel-wise pattern of response for each of the 8 conditions in the first half of the session with each of the 8 conditions in the second half. The red-green 3D glasses were flipped halfway through the session. **C**, Matrices illustrate hypothetical correlation patterns for pure X, Y, and Z location information. For each dimension, we transformed the correlations to z-scores and quantified the amount of location information by subtracting the difference between same-location comparisons (black) and different-location comparisons (white).

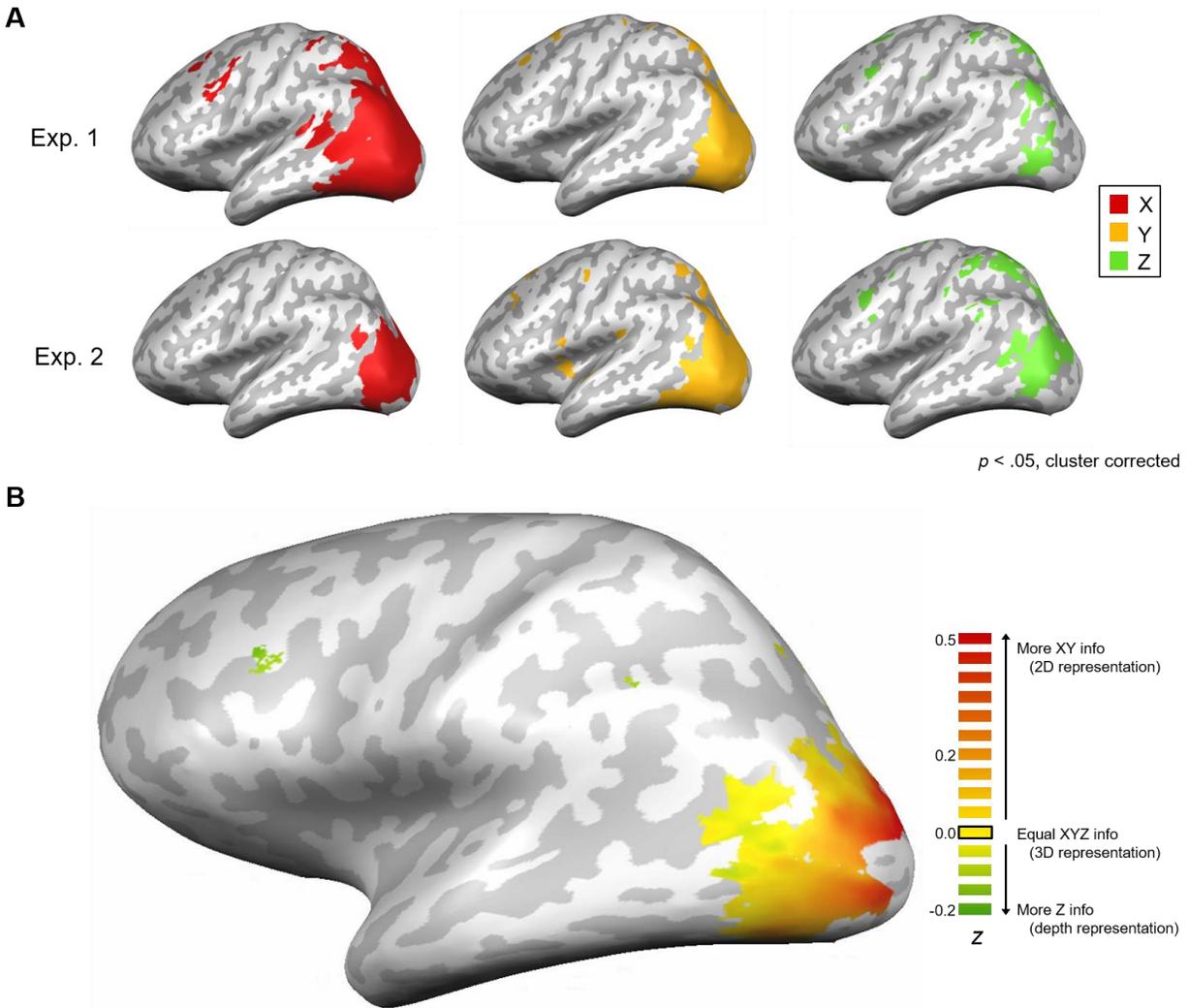


Figure 2. A, Results from the searchlight analysis projected onto inflated brains. Maps of significant X location information (red), Y location information (yellow), and Z location information (green) averaged across subjects for Experiment 1 (N=11) and 2 (N=12). All maps were thresholded at $p < .05$, cluster-corrected. **B**, Hybrid searchlight map of 2D versus 3D spatial representations, averaged across subjects for Experiment 2 (N=12). Map shows the difference in magnitude between 2D information (averaged X and Y information from Fig. 2) and Z information at each voxel. Data are thresholded to show only voxels with significant ($p < 0.05$) information for either XY and/or Z maps.

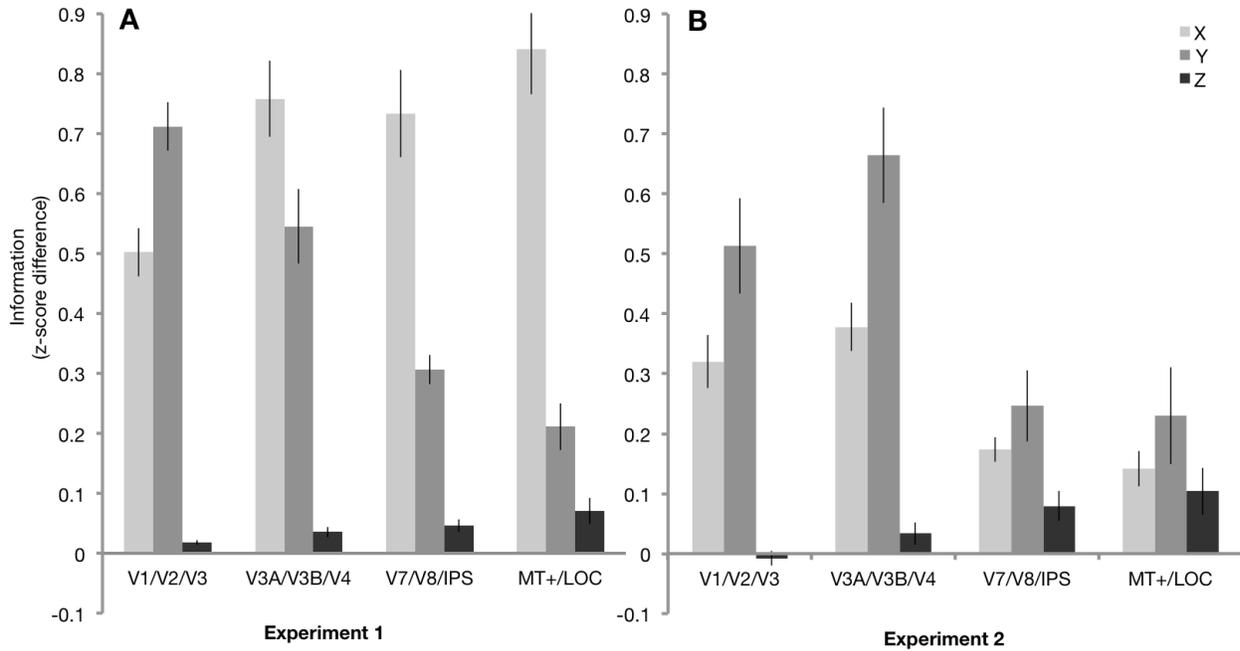


Figure 3. A, Average X, Y, and Z location information within each ROI group for Experiment 1 (N=11) and **B**, Experiment 2 (N=12). Experiment 2 data is shown for the left (contralateral) hemisphere only; data from individual ROIs, including the ipsilateral right hemisphere, are in Figures S4 & S5. Error bars represent SEM.

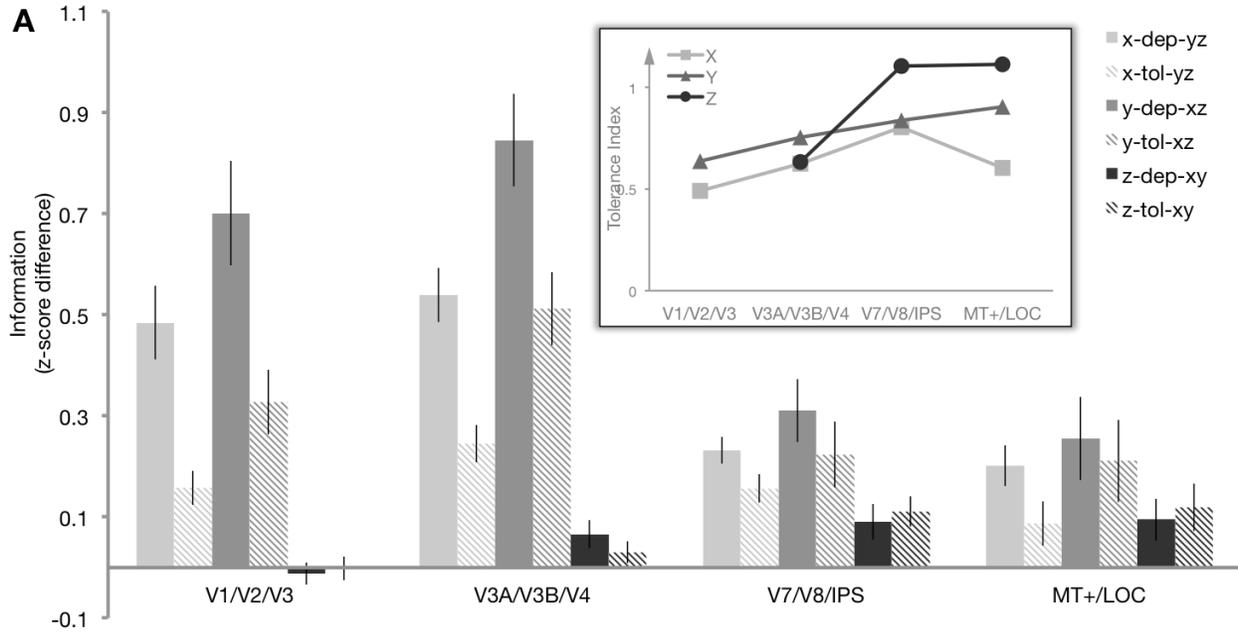


Figure 4. A, Tolerance analysis. Average X, Y, and Z location information within each ROI group for Experiment 2 (N=12), separated into comparisons where the other two dimensions were the same (dependent information, “dep”) or different (tolerant information, “tol”). Note that each bar is a difference in correlations (as in Fig. 3), but calculated for a different subset of cells in the matrix (see Methods). The inset shows the tolerance index for X, Y, and Z location information, calculated as $1 - [(dependent - tolerant) / (dependent + tolerant)]$. Tolerance index was only calculated when there was significant information that could be decoded (i.e., not for Z in early visual areas). A larger index (close to 1) means that the location information for the selected dimension was highly tolerant; a smaller index (close to 0) indicates the location information was highly dependent on the other dimensions. Error bars represent SEM. Data from ungrouped ROIs are shown in Figure S6.

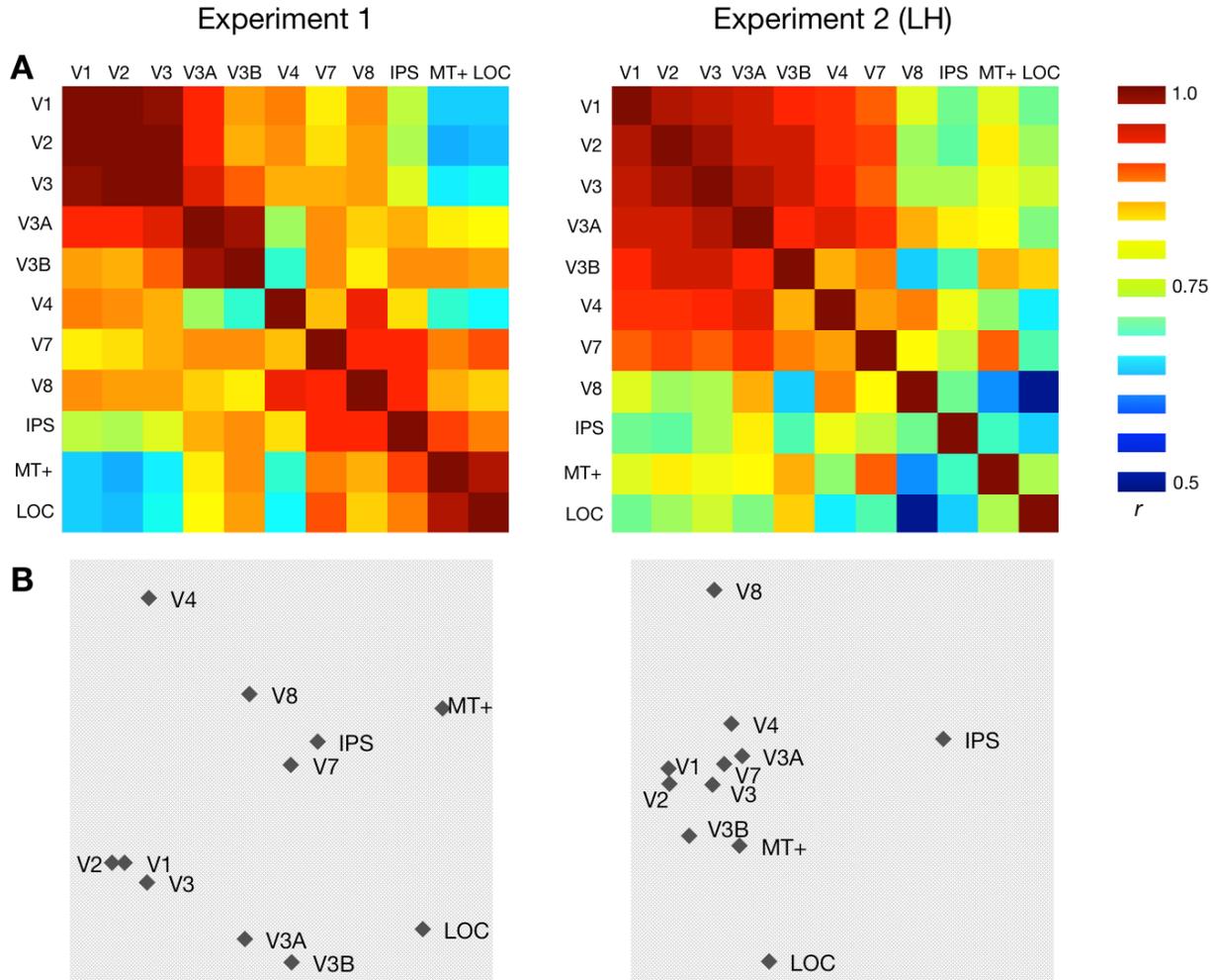


Figure 5. Representational similarity across ROIs. **(a)** Representational Similarity Matrices for Experiment 1 (N=11) and Experiment 2 (N=12) for the 11 ROIs. Each cell represents the correlation between MVPA patterns for a pair of ROIs (symmetric across diagonal). For Experiment 2 only the contralateral (left hemisphere) ROIs were used. **(b)** Multidimensional Scaling visualizations for each experiment. Inter-point distances were calculated using ROI dissimilarity ($1 - r$) matrices, and plotted along the two most informative dimensions.