Decoding 3D spatial location across saccades in human visual cortex

Xiaoli Zhang[1], Christopher M Jones[1], & Julie D Golomb[1]

[1] Department of Psychology, the Ohio State University. 1835 Neil Ave, Columbus, OH 43210.

Correspondence should be addressed to:

Xiaoli Zhang, zhang.4734@osu.edu

Julie D. Golomb, golomb.9@osu.edu

Abstract

Visual signals are initially processed as two-dimensional images on our retina. In daily life, we make frequent eye movements and consequently the 2D retinal inputs constantly change. In addition, to perceive a 3D world, depth information needs to be reconstructed, using cues such as the binocular disparity between the 2D retinal images from both eyes. How do saccades influence the brain representation of 3D spatial locations? In an fMRI scanner, while wearing red-green anaglyph glasses to facilitate 3D perception, participants passively viewed a random dot patch that stimulated one of four 3D locations in each 16-second block. Each location was defined by its 2D position (above or below screen center; vertical information), and its depth position (in front of or behind central screen plane). We compared the amount of 2D and depth information (using multivariate pattern analysis) for no-saccade blocks (in which participants maintained stationary fixation) compared to saccade blocks (a series of guided saccades). In saccade blocks, we could decode vertical and depth information to a similar extent as in no-saccade blocks, despite the retinal changes in horizontal position induced by the saccades. Strikingly, no-saccade blocks exhibited a strong dependence on fixation position: little vertical or depth information could be decoded across blocks with different fixation positions in any visual areas during stable fixation. In contrast, on saccade blocks, both vertical and depth information were tolerant of changes in fixation position. The findings suggest that representations of 3D spatial locations may become more tolerant of fixation positions during “dynamic” saccades, perhaps due to active remapping which may encourage more stable representations of the world.

Keywords: depth, spatial representation, eye movements, fMRI

Introduction

Visual inputs are initially processed on the retina in two-dimensional, eye-centered coordinates, which provides direct coding of 2D spatial locations from the very early stage of visual processing. But our perceptual experience is of a 3D world, and one that is stable across eye movements. We move our eyes around frequently, on the order of a few times per second. As a result, the retinal positions of objects projected to our eyes also change frequently, creating great challenges for visual cognition, and an extra layer of complexity for 3D perception.

Many studies have explored different aspects of this challenge. For example, 2D perception across saccades is known to be disrupted in multiple ways. Objects flashed around the time of a saccade tend to be systematically misperceived as closer to the saccade endpoint than they actually are, a phenomenon called compression of space (Awater & Lappe, 2006; Ross et al., 1997, 2001). Localization of a foveal target is also worse when performed after a saccade compared to without a saccade, and responses tend to be biased towards the former fixation location (Zhang & Golomb, 2018). Furthermore, saccades can interfere with spatial attention (Golomb et al., 2008; Rolfs et al., 2011), memory (Golomb & Kanwisher, 2012a; Irwin, 1991; Prime et al., 2011), and feature perception, including an effect where the features of objects at two different locations can be mixed (Dowd & Golomb, 2019; Golomb et al., 2014). An important question in vision research has been how our brain compensates for the disturbance from executing saccades and maintains stability, via remapping and other mechanisms (reviewed in Golomb, 2019; Melcher, 2011; Wurtz, 2008).

Other studies have explored the mechanisms of depth perception. Our retinal images are only two-dimensional, which means that in order to perceive accurately the three-dimensional world, we need to reconstruct the third dimension, depth, from the 2D retinal inputs. Depth can

be reconstructed from many visual cues, such as size, perspective, shading, motion parallax, and binocular disparity (Howard, 2012). Among these cues, binocular disparity (small horizontal differences in an object's projected location on the two eyes) is particularly effective (Howard & Rogers, 2012). Instead of perceiving two images of the same stimulus, our brain is able to fuse the two images and perceive how far the stimulus is based on how separated the retinal images are (i.e., stereoscopic vision).

In terms of neural representations, it is widely shown that 2D spatial information is represented throughout visual cortex and beyond (Arcaro et al., 2009; DiCarlo & Maunsell, 2003; Engel et al., 1994; Maunsell & Newsome, 1987; Saygin & Sereno, 2008; Sereno et al., 1995; Silver & Kastner, 2009). More controversial are the reference frames of 2D spatial representations across saccades. Some studies have shown that these representations are primarily coded in retinotopic (eye-centered) coordinates, with no evidence for explicit spatiotopic (world-centered) representations (Gardner et al., 2008; Golomb & Kanwisher, 2012b; Zopf et al., 2018). However, other studies have shown some evidence for spatiotopic updating or adaptation across saccades (Crespi et al., 2011; d'Avossa et al., 2007; Fairhall et al., 2017; Golomb et al., 2010; Zimmermann et al., 2016).

Studies examining the neural representations of depth perception have typically focused on sensitivity to binocular disparity (Cumming & Parker, 1997; Finlayson et al., 2017; Henderson et al., 2019; Neri et al., 2004; Preston et al., 2008), and other depth cues (Amit et al., 2012; Lescroart et al., 2015; Nag et al., 2019; Persichetti & Dilks, 2016). A few recent studies have attempted to explore the nature of *3D* spatial representations in the human brain by varying both 2D location and position in depth: A study from our lab revealed that multivoxel pattern information about position-in-depth increases along the visual hierarchy while information about

2D location (horizontal and vertical locations) decreases; that is, representations of spatial locations transition from 2D dominant in early visual areas to balanced 3D in later visual areas (Finlayson et al., 2017). Another study using inverted encoding models similarly recovered representations of both 2D and depth position in visual and parietal areas, particularly in V3A (Henderson et al., 2019).

However, no studies have investigated how these 3D spatial representations are affected by eye movements. Behaviorally, there is some evidence that executing saccades creates challenges for depth processing, which may not be surprising given that reconstructing depth from binocular disparity relies on precise retinal position information, and retinal position changes with each eye movement. Horizontal eye movements in particular can interfere with our processing of depth, e.g., biasing the depth perception of stimuli which flash around the time of a saccade to be closer (Teichert et al., 2008), similar to 2D mislocalization, and impairing memory-guided reaching in depth (Van Pelt & Medendorp, 2008). Interestingly, however, there is also evidence that self-generated motor actions, including eye movements, can *enhance* our perception of 3D space (Wexler & van Boxtel, 2005).

In the current study, we examine how 3D spatial representations in human visual cortex are influenced by saccades. We modified the fMRI paradigm from Finlayson et al. (2017), where participants viewed 3D stimuli in the scanner while wearing red-green anaglyph glasses. We used multivariate pattern analysis (MVPA) to investigate the representations of 2D (vertical; y-axis) location information and depth (z-axis) location information in different visual regions, and compared the representations of 3D location during sustained fixation and dynamic saccade blocks.

Methods

Participants. 12 right-handed subjects participated in the study and were included in the analyses (6 females, 6 males, mean age 21.42, range 19-34). Three additional subjects were also scanned, but the data were excluded due to excessive amount of head motion, failure to perform the dot-dimming task, and radiologist-detected neuroanatomical abnormality, respectively. All subjects included reported normal or corrected to normal vision, and normal color and depth perception. All gave informed consent and were pre-screened for MRI eligibility. The study protocol was approved by the Ohio State University Biomedical Sciences Institutional Review Board. Before the scanning session, each participant went through a series of pre-screening behavior tasks to assess their depth perception (data not shown; all participants scanned exhibited normal depth perception).

fMRI acquisition. This study was done at the OSU Center for Cognitive and Behavioral Brain Imaging with a Siemens Prisma 3T MRI scanner using a 32-channel phase array receiver head coil. Functional data were acquired using a T2-weighted gradient-echo sequence with multiband factor of 3 (TR=2000ms, TE=28ms, flip angle 72°). The slice coverage was oriented parallel to the AC-PC plane and was placed to contain full coverage of cerebrum (72 slices, 2×2×2mm voxel). We also collected a high-resolution MPRAGE anatomical scan at $1mm^3$ resolution for each participant. Each participant was scanned in one two-hour session.

Stimuli and task. The main stimuli and task in the scanner were modified from Finlayson et al. (2017), Experiment 1. We used dynamic random dot stimuli (RDS) to stimulate different 3D

locations in the participants' visual field (Figure 1). We used red/green anaglyph glasses with Psychtoolbox's stereo-mode to achieve depth perception of the RDS stimuli from binocular disparity. The stimuli were small patches (3.3° square) of dynamic RDS located above or below (vertical position) and in front of or behind (depth position) the screen center. The stimulus locations were centered ± 2.7° from the screen center along the vertical dimension, and ± 10 arc min along the depth dimension. A black empty circle (0.06° radius) inside a white dot (0.13° radius) was used as the fixation point. The fixation point could appear in one of two positions: either left or right of the screen center, ± 2.7°, centered vertically and in depth. On half of the blocks (no-saccade blocks), the fixation dot remained in the same position for the entire 16s block. On the other half (saccade blocks), the fixation dot moved back and forth between the two fixation locations every 2 seconds.

To encourage perception of a 3D space, we used a static RDS background field (10.70° square) at the central depth plane of the screen, consisting of light and dark gray dots on a mid-gray background (8 dots/deg^2, 37% contrast). In addition, we used ground and ceiling lineframes (12.8° × 1.5°) above and below this background RDS, respectively, each spanning ± 13 arc min in front and behind the fixation depth plane. Similar to Finlayson et al., 2017, 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).

We employed a block-design, with 8 different stimulation conditions determined by 4 stimulus location conditions and 2 saccade conditions (i.e., Fix-up-front, Fix-up-back, Fix-down-front, Fix-down-back, Sac-up-front, Sac-up-back, Sac-down-front, Sac-down-back). Each run included 16 stimulus blocks, two blocks per condition, to maintain the same power as in Finlayson et al., 2017. We arranged the order of block conditions in a pseudo-random fashion

(i.e., balanced Latin square design) for each run. We added three fixation baseline blocks in each run, where there were no RDS stimuli presented. The fixation blocks were added at the beginning (Block 1), middle (Block 10), and end (Block 19) of each run. All blocks lasted 16s, with a 2s inter-block interval, and each run lasted 360s. Each subject completed 8 runs of the task.

In all conditions, participants were instructed to fixate on the fixation dot and perform a dot dimming task at fixation, detecting when the black empty circle was filled in to be a solid black circle. When the fixation changed its location, participants were instructed to move their eyes to follow it. The fixation position (i.e., left and right) for each no-saccade block and the initial fixation position (determining the saccade direction) for each saccade block was selected as follows: In each run, the sequence of the fixation positions was determined so that there were no inter-block eye movements (i.e., the final fixation position of the previous block was the same as the initial fixation position of the current block). This was important to ensure that saccades were only executed during saccade blocks. The sequence was further constrained to ensure that both left and right fixation positions were equally likely across the 8 conditions over the course of the experiment. The no-saccade conditions were balanced such that for each of the four stimulus location conditions, there was one block per run with left fixation position and one block per run with right fixation position. Because each saccade block involved a repetitive sequence of 8 alternating saccades, an equal number of leftward and rightward saccades were always included in each saccade block. However, for saccade blocks, the distribution of initial fixation positions could not be fully counterbalanced within each run, given the constraint above. E.g., within a given run, the two blocks of the “Saccade task: up-front stimulus location” condition might have both been assigned the left initial fixation position; but across the total 8

runs, an equal number of both left and right initial fixation position blocks were included for each condition.

Participants wore red/green anaglyph glasses to view the 3D stimuli, and were instructed to flip the direction of the glasses midway through the experiment (after four of eight runs), in order to control for low-level stimulus differences in the MVPA analysis due to the color presented to each eye, per Finlayson et al. (2017). When the glasses were flipped, the stimulus code adapted to reflect the current glasses direction and stimulate the correct depth location (i.e., front or back).

All stimuli were generated with Psychtoolbox (Brainard, 1997) in Matlab (MathWorks). Stimuli were displayed with a 3-chips DLP projector onto a screen in the rear of the scanner (resolution 1280×1024 at 60Hz). Participants viewed from a distance of 74cm via a mirror above attached to the head coil.

Eye tracking. Fixation positions were recorded inside the scanner throughout the experiment, using an MRI-compatible Eyelink remote eye-tracker at 500 Hz. Eye position data were used to ensure the participants kept their eyes on the fixation point and made eye movements following the fixation change. Due to the red-green anaglyph glasses interfering reflection, the eye-tracking calibration was not always reliable for all subjects. In circumstances where reliable eye position data were not able to be recorded, the experimenters could observe the subject's eye through the camera video and/or use the behavior performance of the dot dimming task to ensure that the participants were making eye movements as intended.

fMRI preprocessing and analyses. The fMRI data were preprocessed with Brain Voyager QX (Brain Innovation). All functional data were corrected for slice acquisition time and head motion and temporally filtered. Runs with abrupt motion greater than 1mm were discarded from later analyses. Spatial smoothing of 4mm FWHM was performed on the preprocessed data for univariate analyses, but not for multivariate (MVPA) analyses. Data of each participant were normalized into Talairach space (Talairach & Tournoux, 1988). We used FreeSurfer to segment the white matter / gray matter boundaries for each participant's anatomical scan, to inflate and flatten each hemisphere into cortical surface space.

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. More details are given in the multivariate pattern analysis (MVPA) section. All GLM data were exported to Matlab using Brain Voyager's BVQXtools Matlab toolbox, and all subsequent analyses were done using custom code in Matlab.

Functional localizers and Regions of Interest (ROIs). In addition to the main task, each participant also completed 3 runs of functional localizers and 4 runs of 2D retinotopic mapping. We identified a priori ROIs similar to Finlayson et al., 2017:

We defined retinotopic areas V1, V2, V3, V3A, V3B, V4, V7, and V8 using a standard retinotopic mapping paradigm with a rotating wedge with high-contrast radial checkerboard patterns (Sereno et al., 1995). The 60° wedge stimulus covered eccentricity from 1.6° to 16° and flickered at 4 Hz. It was rotated for 7 cycles with a period of 24s per cycle in each run, either clockwise or counterclockwise. Participants' task was to fixate at the center fixation on the screen and press the button when the fixation dot changed color. After preprocessing, the brain

data of retinotopic mapping runs were analyzed in custom Matlab code and were projected onto the flattened cortical surface maps in Brain Voyager, and boundaries between the retinotopic areas were delineated.

For each individual participant, we also defined the object-selective LOC (2 localizer runs) and motion-sensitive MT+ (1 localizer run). The LOC localizer task included blocks of grayscale real-world objects and scrambled objects (12° × 12°) 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 additionally localized a visually sensitive intraparietal sulcus (IPS) ROI using data from the LOC localizer task (All > Fixation contrast) in conjunction with anatomical landmarks.

To compare with findings in Finlayson et al., 2017, we grouped the ROIs in a similar way, according to 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 LOC and MT+. For our main analyses, MVPA information was calculated for individual ROIs and subjects and then averaged across the ROIs in each group. Analyses with all individual ROIs are shown in the supplemental materials as well as whole-brain searchlight MVPA.

Multivoxel pattern analyses (MVPA). Multivoxel pattern analyses (MVPA) following the split-half correlation method (Haxby et al., 2001) were performed for ROI-based as well as whole-brain (searchlight) analyses. Our main approach of quantifying vertical and depth location information was similar to Finlayson et al. (2017).

For our primary analyses, we compared vertical and depth information in the no-saccade versus saccade blocks. We conducted MVPA on the 8 main conditions (4 stimulus location conditions × 2 saccade conditions). We split the data into two halves based on the direction of the anaglyph glasses, correlating multivoxel patterns of activity for each of the 8 conditions in the first 4 runs (RG runs, red color over the left eye) with each of the 8 conditions in the last 4 runs (GR runs, green color over the left eye). We ran GLMs with the 8 conditions as our 8 regressors of interest for each split-half dataset. The beta weights for each voxel were normalized (within each dataset) by subtracting the mean response across all conditions for that voxel from the responses to individual conditions. Next, the voxel-wise response patterns for each of the 8 conditions in the RG runs were correlated with each of the 8 conditions in the GR runs, generating an 8-by-8 correlation matrix (Figure 2A). The correlations were converted to z-scores using Fisher's r-to-z transform.

We then quantified the amount of vertical (Y) and depth (Z) location information contained within each ROI for each subject, as follows. First, the cells in the correlation matrix were characterized according to whether they reflected the same or different Y location, same or different Z location, and same or different saccade condition (no-saccade or saccade). For example, the Sac-up-back(RG) × Sac-down-back(GR) correlation would be characterized as same saccade condition, different Y, same Z (1 0 1). Then, for each type of information, we averaged across all of the "same" cells for that type of information, and all of the "different"

cells (Figure 2A), and the “same” minus “different” correlation difference was taken as a measure of the amount of “information” about that property. E.g., y-information was quantified as the difference in correlation between all conditions that shared the same Y position (- 1 -) versus differed in Y 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 voxel-wise response pattern should be more similar for two conditions that share the same location than differ in location. Y and Z information were calculated in four ways: for all blocks (ignoring saccade information; Figure 2A Matrix 1), within no-saccade blocks only (Figure 2B Matrix 2), within saccade blocks only (Figure 2B Matrix 3), and cross-decoded *between* no-saccade and saccade blocks (Figure 2B Matrix 4). All analyses were performed within each ROI and subject, and then averaged into the ROI groups where applicable. Standard frequentist within-subject statistics were performed comparing whether the different types of information were significantly different from zero and/or each other.

We performed the above analyses on our a priori ROIs, as well as MVPA searchlight analyses (Kriegeskorte et al., 2006) to search through the entire cerebrum for voxel clusters showing significant Y and Z location information. Searchlight results are shown in the supplemental materials.

As a second stage of analysis, we broke down the conditions further to compare within- versus across-fixation location information in no-saccade blocks, and within- versus across-saccade-direction location information in saccade blocks. For each of these analyses we used different GLMs in order to maximize power for the comparisons of interest. The main reason for different GLMs is that no-saccade fixation locations were perfectly balanced within each run, but saccade direction was only balanced across runs, as described earlier.

For the no-saccade breakdown, we analyzed left and right fixation position blocks as separate conditions, running the split-half GLMs with 12 conditions: the 4 stimulus location conditions for each of left- and right-fixation no-saccade conditions (FixLeft-up-front, FixLeft-up-back, FixLeft-down-front, FixLeft-down-back, FixRight-up-front, FixRight-up-back, FixRight-down-front, FixRight-down-back), plus the original 4 stimulus location saccade conditions (modeled in the GLM but ignored in this analysis). Similar to above, we generated correlation matrices between the beta weights of the 8 no-saccade conditions in the RG-runs GLM and the 8 no-saccade conditions in the GR-runs GLM for each ROI (Figure 2B top panel). Y and Z location information were calculated across no-saccade blocks with the same fixation position ("no-saccade same fixation"; Figure 2B Matrix 2a), and across (*between)* no-saccade blocks with different fixation positions ("no-saccade different fixations"; Figure 2B Matrix 2b).

For the saccade direction breakdown, we separated saccade blocks with left and right initial fixation positions into separate conditions, so that there were total 12 conditions: 4 stimulus location conditions for each of left- and right-initial-fixation saccade conditions (SacLeft-up-front, SacLeft-up-back, SacLeft-down-front, SacLeft-down-back, SacRight-up-front, SacRight-up-back, SacRight-down-front, SacRight-down-back), plus the original 4 stimulus location no-saccade conditions (modeled in the GLM but ignored in this analysis). However, because only 4 of the 8 saccade conditions were represented in each individual run, we had to adopt a different split-half approach. We ran GLMs for each run separately, each containing 8 conditions (4 no-saccade conditions ignoring fixation position, and 4 saccade conditions labeled with left or right initial fixations). For each of the 8 saccade conditions (4 stimulus location conditions × 2 initial fixation or saccade direction conditions) for each subject, we then randomly split all the runs containing this condition into two halves (regardless of RG or

GR) and averaged the beta weights across the runs in each half, to create the two split-half voxel-wise patterns for each condition and subject (normalized for mean voxel activation as above within each individual run). These voxel-wise patterns were used for the 8-by-8 correlation matrices for each ROI for each subject (Figure 2B bottom panel). We then calculated Y and Z information across saccade blocks with the same saccade direction ("saccade same direction"; Figure 2B Matrix 3a) and across saccade blocks with different saccade directions ("saccade different directions"; Figure 2B Matrix 3b). We performed 100 iterations of the randomly splitting procedure, and the information results for these 100 iterations were averaged for each subject.

Multivariate pattern time-course analyses. In order to explore the dynamic representations of spatial locations in both no-saccade and saccade blocks, we performed MVPA time-course analyses using finite impulse response (FIR) GLMs with 10 timepoints; timepoint zero (TP0) corresponds to the start of each block (i.e., the onset of the random dot stimulus). To directly compare the time-courses of no-saccade conditions (left vs right fixations) and saccade conditions (left vs right initial fixations) in the same analysis, we ran a GLM including all 16 conditions as regressors of interest. For each run we ran a GLM with 12 regressors of interest corresponding to the 12 conditions included in that run: 8 no-saccade conditions (4 stimulus locations for each of left- and right-fixation conditions), and 4 saccade conditions (4 stimulus locations labeled with either left or right initial fixation or saccade directions). Across 8 runs for each subject, there were a total of 16 conditions (4 stimulus locations × 2 no-saccade fixation locations + 4 stimulus locations × 2 saccade direction conditions). For each subject and condition, we randomly split the dataset into halves and calculated averaged beta weights from

the split-half datasets for each of the 16 conditions. After generating the 16-by-16 correlation matrices for each timepoint, we calculated Y and Z information as well as fixation position / saccade direction information at each point along the time course, for each ROI. This random splitting procedure was repeated with a similar iteration approach as above, and the MVPA results were averaged across the 100 iterations for each subject.

Results

Vertical and Depth location information in visual ROIs

Similar to Finlayson et al. (2017), we calculated the amount of vertical and depth location MVPA information for each pre-defined visual ROI and group, for each type of analysis. To quantify the main effects and interactions, we submitted the data to 2 (location type: Y and Z) × 4 (ROI group) repeated measure ANOVAs. For the analysis collapsing across both saccade and no-saccade blocks (Figure 3A), there was a significant main effect of location type, $F_{1,11}$=35.51, p<.001, η_p^2=.764, indicating greater information about vertical than depth locations across all visual ROIs. Furthermore, we found a significant main effect of ROI group ($F_{1.416,\ 15.580}$=15.71, p<.001, η_p^2=.588), and a significant interaction between location type and ROI group ($F_{1.299,14.286}$=18.89, p<.001, η_p^2=.632). Post-hoc one-way ANOVAs showed that vertical location information decreased along the visual hierarchy, $F_{1.329,14.621}$=17.76, p<.001, η_p^2=.618, whereas depth location information increased, $F_{1.819,20.008}$=5.382, p=.015, η_p^2=.329. Both vertical and depth information were significant in later visual areas (V7/V8/IPS) and category-selective areas (LOC/MT+); post-hoc t-tests for both Y and Z information in these two groups of ROIs showed t's≥2.804, p's≤.017, Cohen's d's≥0.809. These results replicate the findings in Finlayson et al. (2017).

Our primary research question was how 3D location information is represented in dynamic saccade blocks, compared to in no-saccade blocks. We thus separately calculated location information in no-saccade and saccade blocks, and performed a 2 (saccade condition: saccade and no-saccade) × 2 (location type: Y and Z) × 4 (ROI group) three-way repeated measures ANOVA (Figure 3B). We again found a significant main effect of location type ($F_{1,11}$=33.353, p<.001, η_p^2=.752), a significant main effect of ROI group ($F_{1.528,16.804}$=15.661, p<.001, η_p^2=.587), and a significant interaction between location type and ROI group ($F_{1.326,14.584}$=17.645, p<.001, η_p^2=.616). Most importantly, there was no significant main effect of saccade condition in the three-way ANOVA, $F_{1,11}$=1.046, p=.328, η_p^2=.087, nor were any interactions with saccade condition significant (all Fs≤1.394, ps≥.270, η_p^2s≤.112). Post-hoc t-tests confirmed that in saccade blocks, both vertical and depth information were significant in later areas and category-selective areas (t's≥2.417, p's≤.034, Cohen's d's≥0.698). This result revealed that both vertical (2D) and depth information in these later areas was preserved in saccade blocks, such that making dynamic saccades did not seem to impair the decodability of location information at all, compared to in no-saccade blocks.

Location representations shared between saccade blocks and no-saccade blocks

Although we could decode both Y and Z information in saccade blocks in later visual areas without impairment, a further question is whether similar brain patterns for location were shared between no-saccade and saccade blocks, or if distinct brain patterns were used to represent spatial location separately during fixation and across saccades. To answer this question, we attempted to cross-decode location information in the two contexts, by correlating the brain patterns *between* no-saccade and saccade blocks (Figure 2B Matrix 4).

As shown in Figure 3C, the cross-block MVPA results were similar to the results within no-saccade blocks or saccade blocks alone: There was a significant main effect of location type ($F_{1,11}$=34.75, p<.001, η_p^2=.760), a significant main effect of ROI group ($F_{1.361,14.975}$=15.06, p<.001, η_p^2=.578), and a significant interaction ($F_{1.311,14.416}$=19.35, p<.001, η_p^2=.638). Vertical location information could be significantly decoded in all four ROI groups (t's≥3.997, p's≤.002, Cohen's d's≥1.154), and depth location information could be significantly decoded in LOC/MT+ (t_{11}=4.129, p=.002, Cohen's d =1.192). This suggests that saccade and no-saccade blocks share similar activation patterns for a stimulus appearing in a given vertical and/or depth location.

Tolerance of location information across fixation positions and saccade directions

We next asked whether vertical and depth location information are dependent on fixation position and/or saccade direction, or whether they are tolerant of these differences. To test if the representations of vertical and depth location are tolerant of changes in fixation position, we separated the no-saccade blocks with left and right fixation positions, and then separately analyzed vertical and depth information within blocks that shared the same fixation position ("same-fixation"), and between blocks of different fixation positions ("different-fixation").

As shown in Figure 4A, we found significant information about both vertical and depth location from the MVPA correlations across no-saccade blocks sharing the same fixation position, but not across blocks with different fixation positions. We ran a 2 (fixation position: same-fixation and different-fixation) × 2 (location type: Y and Z) × 4 (ROI group) three-way repeated measures ANOVA. In addition to the significant main effects of location type and ROI group, there was also a significant main effect of fixation position, as well as significant interactions between fixation position and location type, fixation position and ROI group, and

among all three variables (all Fs≥10.97, ps<.001, η_p^2s≥.499). This suggests that the representations of both vertical and depth location were largely dependent on fixation position in no-saccade blocks.

On the other hand, when we separated the saccade blocks with left and right initial fixation positions (i.e., different saccade directions), we found a different pattern. Using a similar approach to the above analyses for no-saccade blocks (replacing same/different fixation location with same/different saccade direction), we found the standard significant main effects of location type and ROI group, as well as a significant interaction between location type and ROI group (Fs≥18.922, ps<.001, η_p^2s≥.632). However, here there was no significant main effect of saccade direction ($F_{1,11}$=0.072, p=.793, η_p^2=.007); overall information about both Y and Z was comparably maintained during saccade blocks, even across blocks with different saccade directions (i.e. different fixation locations at each point in the block) (Figure 4B). There was a significant interaction between saccade direction and location type, $F_{1,11}$=22.277, p<.001, η_p^2=.669, but all other interactions were not significant, Fs≤1.816, ps≥.197, η_p^2s≤.142. The post-hoc t-tests revealed that z information was significant in all ROI groups in the different-direction condition, t's≥2.525, p's≤.028, Cohen's d's≥0.729. However, z information in the same-direction condition was not significantly above zero in these regions, t's≤2.121, p's≥.057, Cohen's d's≤0.612. Comparing z information between same-direction and different-direction conditions, there was significant difference only in early and later visual ROIs (t's≥2.814, p's≤.017, Cohen's d's≥0.812), but not in intermediate and category-selective ROIs (t's≤0.453 p's≥.660, Cohen's d's≤0.131). The less reliable z information may be due to that fact that this analysis was substantially less powered than the earlier analyses, due to (1) looking at only a subset of the whole dataset (i.e., 1/8 of the power compared to Figure 3A; as shown in Figure 2 correlation

matrices) and (2) single-run GLMs required for the saccade direction breakdown may have introduced more noise compared to the analogous no-saccade breakdown. Perhaps as a result of the decreased power, the amount of both vertical and depth location information was decreased in scale compared to Figure 3. Nonetheless, the data present a striking result, even for the more robust vertical location information: information about the stimulus' vertical location was tolerant of saccade direction and thus fixation position in the dynamic saccade blocks, but was not tolerant of fixation position in the static no-saccade blocks.

Timepoint-by-timepoint tolerance analysis

The above analyses revealed that vertical and depth location information could be decoded during saccade blocks, and appeared largely tolerant of saccade direction. This is particularly interesting in comparison to the lack of tolerance to fixation position found in the no-saccade blocks. The different saccade directions by definition involve different fixation locations; in other words, saccade blocks could be considered as a series of alternating short periods (2 s) with left and right fixation positions. When comparing the two different saccade directions, at any given point in the block, the fixation position is different for a sacLR versus sacRL block. The fact that stimulus location information was tolerant to changes in fixation position in the saccade blocks but not the no-saccade blocks is a striking result. One explanation we entertain later in the discussion is that this difference could be due to the dynamic, active context of the saccade blocks; i.e. that the saccades trigger a more tolerant representation. But there are less interesting possibilities as well that could be artifacts of the block design. The GLMs used in the above analyses modeled the whole 16 second block as a single event, which could have effectively blended, or combined, activity across the temporal sequence of left and

right fixations. In other words, it's possible that the apparent tolerance across saccade direction could have been an artifact of the activation patterns containing location information from *both* left and right fixation positions during saccade blocks, regardless of saccade direction.

To address this possibility, we also performed an MVPA time-course analysis, conducting the MVPA contrasts reported above (same fixation location, different fixation location, same saccade direction, different saccade direction), but now on activity patterns for each timepoint estimated from FIR GLMs (see methods). Perhaps as a result of this being a more underpowered analysis, the magnitude of depth information was too small to draw reliable conclusions in this timepoint-by-timepoint analysis (supplemental Figure S3 & Tables S2-S3). For the vertical information timecourses (Figure 5), in the early, intermediate, and later visual areas (the regions in which we could best decode vertical location information in the whole-trial MVPA), the timecourse of decoding peaks after several seconds and remains sustained for the block, for three out of the four contrasts. Critically, vertical information during saccade blocks could be successfully decoded on a timepoint by timepoint basis even across blocks with different saccade directions (different fixation positions at each timepoint), but it was not tolerant of different fixation positions during no-saccade blocks.

Discussion

In the current study we compared how representations of 3D spatial information in human visual cortex are influenced by eye movements. First, our results during fixation blocks replicated the previous findings of Finlayson et al., 2017: MVPA information about the 2D (vertical) location of a stimulus decreased along the visual hierarchy whereas depth location information increased, such that 3D location could be decoded from later visual areas. Critically,

we revealed that both vertical and depth location could be decoded to a similar extent in dynamic saccade blocks, where participants made a sequence of back and forth saccades while passively viewing the stimulus. Follow-up analyses separating fixation positions and saccade directions showed that the representations of 3D spatial locations (both vertical and depth) were *dependent* on fixation position in no-saccade blocks, yet exhibited *tolerance* across fixation changes and different saccade directions in saccade blocks. This pattern persisted even when analyzing timepoint by timepoint, at least for the more robust vertical location information.

Dependence on fixation position in no-saccade blocks

One of our more striking results was that we barely found any vertical or depth information tolerant of fixation position in the no-saccade blocks. In the current design, because stimuli always appeared in the horizontal center of the screen, the two different fixation positions by definition resulted in different eye-centered (retinotopic) horizontal locations. Visual regions are known to be organized in retinotopic reference frames (Gardner et al., 2008; Golomb & Kanwisher, 2012b), and moreover, horizontal information may dominate spatial representations, especially when horizontal location is varied across hemifields (Finlayson et al., 2017; Schneider et al., 1993). Because we could not separate these two factors in this experiment, the vertical and depth location information could have been dependent on fixation position and/or horizontal stimulus location. However, Finlayson et al. (2017) *did* vary horizontal stimulus location alongside vertical and depth information in a very similar design (but with fixation position always held at center). In that study, both vertical and depth locations were found to be at least partially tolerant of horizontal location. This comparison raises the intriguing possibility that representations of vertical and depth locations are dependent on (horizontal) fixation/eye

position, but not so much on horizontal stimulus location, even though both manipulations produce the same retinal consequences.

A number of prior studies have reported findings consistent with a role of fixation/eye position in neural processing of space. Neurophysiological studies have shown neurons' "gain fields" – where the neural response depends on the angle of gaze (Andersen & Mountcastle, 1983). In addition, information about fixation position can be found alongside retinotopic stimulus location from fMRI in primate and human visual cortex (Golomb & Kanwisher, 2012b; Lehky et al., 2008; Merriam et al., 2013; Rosenbluth & Allman, 2002; Trotter & Celebrini, 1999), and static fixation positions have been shown to modulate the response amplitude of retinotopic visual stimuli measured in human fMRI (Merriam et al., 2013; Strappini et al., 2015). Our findings reveal a particularly strong modulation by fixation position on the representation of both 2D and depth locations, such that the representations were completely dependent on fixation position. However, it is possible that the combination of fixation position difference and retinotopic / hemifield difference exacerbated the dependence. Regardless, having retinal images "labeled" with the corresponding fixation positions is likely an important mechanism for visual stability (Cohen & Andersen, 2002; Melcher & Morrone, 2015; Wurtz, 2008). That said, the strong dependence on fixation position we found in no-saccade blocks is particularly interesting in light of our other finding, that spatial locations *could* be decoded in dynamic saccade blocks, tolerant of changes in fixation position.

Fixation-tolerant representations of 3D location in saccade blocks

We found that both 2D and depth location information could be decoded in saccade blocks in later visual areas; in fact, we did not observe any degradation of spatial information in

saccade blocks compared to no-saccade blocks, even when decoding across blocks with different saccade directions / fixation patterns. We had predicted that horizontal saccades might disrupt depth representations in particular, since horizontal binocular disparity is a main cue for depth perception (indeed, the only cue manipulated in the present experiment), and there is some behavioral evidence showing that horizontal saccades impair depth processing, such as depth judgments of flashed stimuli (Teichert et al., 2008), and memory-guided reaching in depth (Van Pelt & Medendorp, 2008). However, the fact that we could reliably decode 3D spatial location from brain activity during the saccade blocks doesn't mean there is no disruption. First, our task differed from the behavioral depth tasks in that participants were not required to process the precise depth locations to perform the task; if we had probed more precise depth locations or included an attention to depth component, it's possible there may have been a more subtle decrement. Second, there could still have been some transient disruption triggered by the saccade, but the representations could have quickly remapped / recovered after each saccade. Remapping can occur predictively (Duhamel et al., 1992) and is at most completed within a few hundred milliseconds after the saccade (Golomb et al., 2008); in the current task the stimulus was present the whole time and saccades were spaced with a 2s gap, so there could have be enough time for the representations of spatial locations to remap and re-stabilize, without the BOLD signal capturing this transient change. Regardless, our findings show that representations of both depth and 2D spatial locations can at least be rapidly updated and preserved across saccades, and – critically – that representations of 3D spatial locations across dynamic saccades are not just a simple aggregation of those representations used for sustained fixations.

The role of active, dynamic saccades in spatial stability

Our key finding was a pattern of fixation-position-dependency during stable fixation and fixation-position-tolerance during dynamic saccades, for both 2D vertical (y) and depth (z) spatial dimensions. We speculate that the difference between tolerance in saccade versus no-saccade blocks seems to fall on whether active, frequent, and repetitive eye movements were made, such that representations of 3D spatial location (both 2D and depth) *become* more tolerant of fixation position during – and because of – active, dynamic saccades.

As briefly mentioned in Results section, a potential alternate explanation is one of a temporal blending artifact from the block-design GLM analysis, where the activation pattern on saccade blocks might reflect a combination of left and right fixation patterns, such that the specific saccade pattern / direction becomes meaningless across the block. However, if this were the dominant reason, we would probably expect decoding performance in the saccade blocks to be worse overall compared to no-saccade blocks; given the strong fixation position dependence on no-saccade blocks, if activity patterns were contaminated between left and right fixation periods on saccade trials, we would predict a cost for decoding. Second, the MVPA timecourse analysis argues against this idea. Although our design was not optimized for time-course analyses and individual time points are not independent, we see little evidence that the whole-block results could be solely attributable to blending. Instead, the results seem more consistent with tolerance emerging due to some active processing mechanism across saccades.

The idea of active processing is not new. Self-generated actions are known to influence perception in a variety of ways (Witt, 2011). It has been found that executing or even preparing motor actions can modify perception and representation of 3D space, beyond the direct sensory consequences from the motor actions (reviewed in Wexler & van Boxtel, 2005). In the case of eye movements, the motor act generates a corollary discharge, or efferent feedback, signal that

feeds back to visual areas and is thought to be a key mechanism for remapping and stabilizing perception (Wurtz, 2018). This corollary discharge signal can even precede the motor movement itself, allowing for predictive remapping in 2D (Sommer & Wurtz, 2006) and 3D (Wexler, 2005). Indeed, in a classic study where the eye muscles were paralyzed and eye movements were intended but unable to execute, corollary discharge resulted in false perception of a displacement of the visual scene (Stevens et al., 1976); when the thalamus is lesioned, patients may mis-attribute perceptual consequences of oculomotor targeting error to external stimulus changes (Ostendorf et al., 2010). In typical vision, we perceive objects in the world as stationary when we're moving our eyes, but an object displacing the analogous amount on our retinas without an active eye movement would be readily detected. Consistently, an fMRI study showed that scene representations spanning different views can be integrated across eye movements in scene-selective cortex, but not in the condition where eyes were stable and the scene was moved to mimic the retinal changes induced by eye movements (Golomb et al., 2011).

Thus, it seems likely that in our study, the blocks in which participants were actively making saccades triggered additional neural signals not present during the sustained fixation blocks, and this could have contributed to greater integration and stability, resulting in more tolerant representations[1]. What is equally intriguing, though, is the amount of fixation position *dependence* found in the no-saccade blocks, even for spatial dimensions that are not directly

[1] Note that in the current study we did not actually test whether the location representations became more stable due to dynamic saccades or dynamic retinal changes. One could imagine a control condition where the fixation was kept stable and the stimulus jumped between left and right sides every 2s. However, this condition would likely produce an inconclusive answer, because as noted earlier, Finlayson et al., 2017 found that vertical and depth location information was partially tolerant to horizontal location during sustained (stable) fixation, so in this case we would expect to see a tolerant representation in the hypothetical control as well.

affected by the change in eye position (i.e. vertical and depth information, across different horizontal fixation positions). The idea of eye position gain fields discussed earlier underlies another theory of visual stability; that spatiotopic (world-centered) neural representations are formed implicitly by the combination of current retinotopic position and current eye position (Melcher & Morrone, 2015; Merriam et al., 2013; Wurtz, 2008), the latter signal coming from proprioceptive information relaying the position of the eye in the orbit (Bridgeman, 1995). An interesting interpretation of our current findings is that there may be different stability-related signals during active vs static perception, such that 3D location is labeled with eye position information during fixation, but this information is integrated in the context of active eye movements. This maps nicely onto the idea that visual stability across saccades incorporates two distinct sources of feedback operating at different timescales: a rapid, predictive corollary discharge signal (triggered by an *active eye movement*), and an oculomotor proprioceptive (*static*) signal that stabilizes more slowly after a saccade (Sun & Goldberg, 2016).

In addition to active remapping mechanisms, we can also consider some other possible mechanisms underlying our observation of 3D location representations becoming tolerant in the context of active, dynamic saccades. These might include factors related to the predictability and expectation of the saccades sequence and/or repetitive retinal changes, as well as questions about whether the increased tolerance could be achieved with a single saccade or is built up over the course of several repeated actions. Follow-up investigation is still needed to further test these possibilities, but all of these possibilities indicate the important role of an active observer, in contrast with a passive viewer.

In sum, our findings highlight the important role of active, dynamic saccades on stabilizing 3D spatial representations in the brain. Even though the horizontal saccades may have

introduced a challenge for preserving accurate stereopsis through binocular disparity, information about stimulus position in depth in human visual cortex was not diminished. If anything, the representations became more robust and tolerant. Given the strong similarity between the depth and 2D (vertical) patterns, our findings could likely be extended to other depth cues and spatial dimensions. However, it seems likely that task settings also matter, where based on task requirements, different processing or different heuristics can be applied. Our study provides support to a potential mechanism, triggered by active saccades, for spatial location processing across eye movements, and indicates that we may represent the visual world more flexibly, based on different situations (e.g., whether sustained fixations or frequent saccades are expected).

References

Amit, E., Mehoudar, E., Trope, Y., & Yovel, G. (2012). Do object-category selective regions in the ventral visual stream represent perceived distance information? *Brain and Cognition*, *80*(2), 201–213. https://doi.org/10.1016/j.bandc.2012.06.006

Andersen, R. A., & Mountcastle, V. B. (1983). The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *3*(3), 532–548.

Arcaro, M. J., McMains, S. A., Singer, B. D., & Kastner, S. (2009). Retinotopic Organization of Human Ventral Visual Cortex. *Journal of Neuroscience*, *29*(34), 10638–10652. https://doi.org/10.1523/JNEUROSCI.2807-09.2009

Awater, H., & Lappe, M. (2006). Mislocalization of Perceived Saccade Target Position Induced by Perisaccadic Visual Stimulation. *Journal of Neuroscience*, *26*(1), 12–20. https://doi.org/10.1523/JNEUROSCI.2407-05.2006

Bridgeman, B. (1995). A review of the role of efference copy in sensory and oculomotor control systems. *Annals of Biomedical Engineering*, *23*(4), 409–422. https://doi.org/10.1007/BF02584441

Cohen, Y. E., & Andersen, R. A. (2002). A common reference frame for movement plans in the posterior parietal cortex. *Nature Reviews Neuroscience*, *3*(7), 553–562. https://doi.org/10.1038/nrn873

Crespi, S., Biagi, L., d'Avossa, G., Burr, D. C., Tosetti, M., & Morrone, M. C. (2011). Spatiotopic Coding of BOLD Signal in Human Visual Cortex Depends on Spatial Attention. *PLOS ONE*, *6*(7), e21661. https://doi.org/10.1371/journal.pone.0021661

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

d'Avossa, G., Tosetti, M., Crespi, S., Biagi, L., Burr, D. C., & Morrone, M. C. (2007). Spatiotopic selectivity of BOLD responses to visual motion in human area MT. *Nature Neuroscience*, *10*(2), 249–255. https://doi.org/10.1038/nn1824

DiCarlo, J. J., & Maunsell, J. H. R. (2003). Anterior Inferotemporal Neurons of Monkeys Engaged in Object Recognition Can be Highly Sensitive to Object Retinal Position. *Journal of Neurophysiology*, *89*(6), 3264–3278. https://doi.org/10.1152/jn.00358.2002

Dowd, E. W., & Golomb, J. D. (2019). The Binding Problem after an eye movement. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-019-01739-y

Duhamel, J.-R., Colby, C. L., & Goldberg, M. E. (1992). The Updating of the Representation of Visual Space in Parietal Cortex by Intended Eye Movements. *Science*, *255*(5040), 90.

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*, *369*(6481), 525–525. https://doi.org/10.1038/369525a0

Fairhall, S. L., Schwarzbach, J., Lingnau, A., Van Koningsbruggen, M. G., & Melcher, D. (2017). Spatiotopic updating across saccades revealed by spatially-specific fMRI adaptation. *NeuroImage*, *147*, 339–345. https://doi.org/10.1016/j.neuroimage.2016.11.071

Finlayson, N. J., Zhang, X., & Golomb, J. D. (2017). Differential patterns of 2D location versus depth decoding along the visual hierarchy. *NeuroImage*, *147*, 507–516. https://doi.org/10.1016/j.neuroimage.2016.12.039

Gardner, J. L., Merriam, E. P., Movshon, J. A., & Heeger, D. J. (2008). Maps of Visual Space in Human Occipital Cortex Are Retinotopic, Not Spatiotopic. *Journal of Neuroscience*, *28*(15), 3988–3999. https://doi.org/10.1523/JNEUROSCI.5476-07.2008

Golomb, J. D. (2019). Remapping locations and features across saccades: A dual-spotlight theory of attentional updating. *Current Opinion in Psychology*, *29*, 211–218. https://doi.org/10.1016/j.copsyc.2019.03.018

Golomb, J. D., Albrecht, A. R., Park, S., & Chun, M. M. (2011). Eye Movements Help Link Different Views in Scene-Selective Cortex. *Cerebral Cortex (New York, NY)*, *21*(9), 2094–2102. https://doi.org/10.1093/cercor/bhq292

Golomb, J. D., Chun, M. M., & Mazer, J. A. (2008). The Native Coordinate System of Spatial Attention Is Retinotopic. *Journal of Neuroscience*, *28*(42), 10654–10662. https://doi.org/10.1523/JNEUROSCI.2525-08.2008

Golomb, J. D., & Kanwisher, N. (2012a). Retinotopic memory is more precise than spatiotopic memory. *Proceedings of the National Academy of Sciences*, *109*(5), 1796–1801. https://doi.org/10.1073/pnas.1113168109

Golomb, J. D., & Kanwisher, N. (2012b). Higher Level Visual Cortex Represents Retinotopic, Not Spatiotopic, Object Location. *Cerebral Cortex*, *22*(12), 2794–2810. https://doi.org/10.1093/cercor/bhr357

Golomb, J. D., L'Heureux, Z. E., & Kanwisher, N. (2014). Feature-Binding Errors After Eye Movements and Shifts of Attention. *Psychological Science*, *25*(5), 1067–1078. https://doi.org/10.1177/0956797614522068

Golomb, J. D., Nguyen-Phuc, A. Y., Mazer, J. A., McCarthy, G., & Chun, M. M. (2010). Attentional Facilitation throughout Human Visual Cortex Lingers in Retinotopic Coordinates after Eye Movements. *Journal of Neuroscience*, *30*(31), 10493–10506. https://doi.org/10.1523/JNEUROSCI.1546-10.2010

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

Henderson, M., Vo, V., Chunharas, C., Sprague, T., & Serences, J. (2019). Multivariate Analysis of BOLD Activation Patterns Recovers Graded Depth Representations in Human Visual and Parietal Cortex. *ENeuro*, *6*(4). https://doi.org/10.1523/ENEURO.0362-18.2019

Howard, I. P. (2012). *Perceiving in depth, Vol. 1: Basic mechanisms*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199764143.001.0001

Howard, I. P., & Rogers, B. J. (2012). *Perceiving in depth, Vol. 2: Stereoscopic vision*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199764150.001.0001

Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology*, *23*(3), 420–456. https://doi.org/10.1016/0010-0285(91)90015-G

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868. https://doi.org/10.1073/pnas.0600244103

Lehky, S. R., Peng, X., McAdams, C. J., & Sereno, A. B. (2008). Spatial Modulation of Primate Inferotemporal Responses by Eye Position. *PLOS ONE*, *3*(10), e3492. https://doi.org/10.1371/journal.pone.0003492

Lescroart, M. D., Stansbury, D. E., & Gallant, J. L. (2015). Fourier power, subjective distance, and object categories all provide plausible models of BOLD responses in scene-selective visual areas. *Frontiers in Computational Neuroscience*, *9*. https://doi.org/10.3389/fncom.2015.00135

Maunsell, J. H., & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience*, *10*, 363–401. https://doi.org/10.1146/annurev.ne.10.030187.002051

Melcher, D. (2011). Visual stability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1564), 468–475. https://doi.org/10.1098/rstb.2010.0277

Melcher, D., & Morrone, M. C. (2015). Nonretinotopic visual processing in the brain. *Visual Neuroscience*, *32*. https://doi.org/10.1017/S095252381500019X

Merriam, E. P., Gardner, J. L., Movshon, J. A., & Heeger, D. J. (2013). Modulation of Visual Responses by Gaze Direction in Human Visual Cortex. *Journal of Neuroscience*, *33*(24), 9879–9889. https://doi.org/10.1523/JNEUROSCI.0500-12.2013

Nag, S., Berman, D., & Golomb, J. D. (2019). Category-selective areas in human visual cortex exhibit preferences for stimulus depth. *NeuroImage*, *196*, 289–301. https://doi.org/10.1016/j.neuroimage.2019.04.025

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

Ostendorf, F., Liebermann, D., & Ploner, C. J. (2010). Human thalamus contributes to perceptual stability across eye movements. *Proceedings of the National Academy of Sciences*, *107*(3), 1229–1234. https://doi.org/10.1073/pnas.0910742107

Persichetti, A. S., & Dilks, D. D. (2016). Perceived egocentric distance sensitivity and invariance across scene-selective cortex. *Cortex*, *77*, 155–163. https://doi.org/10.1016/j.cortex.2016.02.006

Preston, T. J., Li, S., Kourtzi, Z., & Welchman, A. E. (2008). Multivoxel Pattern Selectivity for Perceptually Relevant Binocular Disparities in the Human Brain. *Journal of Neuroscience*, *28*(44), 11315–11327. https://doi.org/10.1523/JNEUROSCI.2728-08.2008

Prime, S. L., Vesia, M., & Crawford, J. D. (2011). Cortical mechanisms for trans-saccadic memory and integration of multiple object features. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1564), 540–553. https://doi.org/10.1098/rstb.2010.0184

Rolfs, M., Jonikaitis, D., Deubel, H., & Cavanagh, P. (2011). Predictive remapping of attention across eye movements. *Nature Neuroscience*, *14*(2), 252–256. https://doi.org/10.1038/nn.2711

Rosenbluth, D., & Allman, J. M. (2002). The Effect of Gaze Angle and Fixation Distance on the Responses of Neurons in V1, V2, and V4. *Neuron*, *33*(1), 143–149. https://doi.org/10.1016/S0896-6273(01)00559-1

Ross, J., Morrone, M. C., & Burr, D. C. (1997). Compression of visual space before saccades. *Nature*, *386*(6625), 598–601. https://doi.org/10.1038/386598a0

Ross, J., Morrone, M. C., Goldberg, M. E., & Burr, D. C. (2001). Changes in visual perception at the time of saccades. *Trends in Neurosciences*, *24*(2), 113–121. https://doi.org/10.1016/S0166-2236(00)01685-4

Saygin, A. P., & Sereno, M. I. (2008). Retinotopy and Attention in Human Occipital, Temporal, Parietal, and Frontal Cortex. *Cerebral Cortex*, *18*(9), 2158–2168. https://doi.org/10.1093/cercor/bhm242

Schneider, W., Noll, D. C., & Cohen, J. D. (1993). Functional topographic mapping of the cortical ribbon in human vision with conventional MRI scanners. *Nature*, *365*(6442), 150–153. https://doi.org/10.1038/365150a0

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., & al, et. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, *268*(5212), 889.

Silver, M. A., & Kastner, S. (2009). Topographic maps in human frontal and parietal cortex. *Trends in Cognitive Sciences*, *13*(11), 488–495. https://doi.org/10.1016/j.tics.2009.08.005

Sommer, M. A., & Wurtz, R. H. (2006). Influence of the thalamus on spatial visual processing in frontal cortex. *Nature*, *444*(7117), 374–377. https://doi.org/10.1038/nature05279

Stevens, J. K., Emerson, R. C., Gerstein, G. L., Kallos, T., Neufeld, G. R., Nichols, C. W., & Rosenquist, A. C. (1976). Paralysis of the awake human: Visual perceptions. *Vision Research*, *16*(1), 93-IN9. https://doi.org/10.1016/0042-6989(76)90082-1

Strappini, F., Pitzalis, S., Snyder, A. Z., McAvoy, M. P., Sereno, M. I., Corbetta, M., & Shulman, G. L. (2015). Eye position modulates retinotopic responses in early visual areas: A bias for

the straight-ahead direction. *Brain Structure and Function*, *220*(5), 2587–2601. https://doi.org/10.1007/s00429-014-0808-7

Sun, L. D., & Goldberg, M. E. (2016). Corollary Discharge and Oculomotor Proprioception: Cortical Mechanisms for Spatially Accurate Vision. *Annual Review of Vision Science*, *2*(1), 61–84. https://doi.org/10.1146/annurev-vision-082114-035407

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain. 3-Dimensional proportional system: An approach to cerebral imaging*. Thieme.

Teichert, T., Klingenhoefer, S., Wachtler, T., & Bremmer, F. (2008). Depth perception during saccades. *Journal of Vision*, *8*(14), 27–27. https://doi.org/10.1167/8.14.27

Trotter, Y., & Celebrini, S. (1999). Gaze direction controls response gain in primary visual-cortex neurons. *Nature*, *398*(6724), 239–242. https://doi.org/10.1038/18444

Van Pelt, S., & Medendorp, W. P. (2008). Updating Target Distance Across Eye Movements in Depth. *Journal of Neurophysiology*, *99*(5), 2281–2290. https://doi.org/10.1152/jn.01281.2007

Wexler, M. (2005). Anticipating the three-dimensional consequences of eye movements. *Proceedings of the National Academy of Sciences*, *102*(4), 1246–1251. https://doi.org/10.1073/pnas.0409241102

Wexler, M., & van Boxtel, J. J. A. (2005). Depth perception by the active observer. *Trends in Cognitive Sciences*, *9*(9), 431–438. https://doi.org/10.1016/j.tics.2005.06.018

Witt, J. K. (2011). Action's Effect on Perception. *Current Directions in Psychological Science*, *20*(3), 201–206. https://doi.org/10.1177/0963721411408770

Wurtz, R. H. (2008). Neuronal mechanisms of visual stability. *Vision Research*, *48*(20), 2070–2089. https://doi.org/10.1016/j.visres.2008.03.021

Wurtz, R. H. (2018). Corollary Discharge Contributions to Perceptual Continuity Across Saccades. *Annual Review of Vision Science*, *4*(1), 215–237. https://doi.org/10.1146/annurev-vision-102016-061207

Zhang, X., & Golomb, J. D. (2018). Target localization after saccades and at fixation: Nontargets both facilitate and bias responses. *Visual Cognition*, *26*(9), 734–752. https://doi.org/10.1080/13506285.2018.1553810

Zimmermann, E., Weidner, R., Abdollahi, R. O., & Fink, G. R. (2016). Spatiotopic Adaptation in Visual Areas. *Journal of Neuroscience*, *36*(37), 9526–9534. https://doi.org/10.1523/JNEUROSCI.0052-16.2016

Zopf, R., Butko, M., Woolgar, A., Williams, M. A., & Rich, A. N. (2018). Representing the location of manipulable objects in shape-selective occipitotemporal cortex: Beyond retinotopic reference frames? *Cortex*, *106*, 132–150. https://doi.org/10.1016/j.cortex.2018.05.009

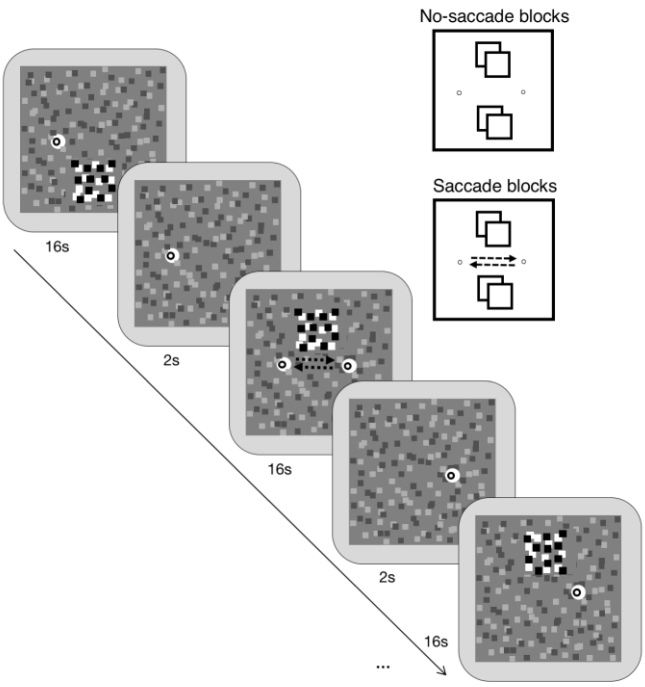

Figure 1 Experiment paradigm. Each block lasted 16s, with 2s inter-block interval. In each block, a dynamic RDS patch stimulated one of four locations, defined by its vertical location (above or below the screen center) and its depth location (in front of or behind the screen depth plane). In half of the blocks, participants kept fixated at the fixation dot at either left or right of the screen center throughout the block; in the other half, participants made repetitive saccades between the left and right, following the alternating fixation dot (as shown in the schematics in the upper right).

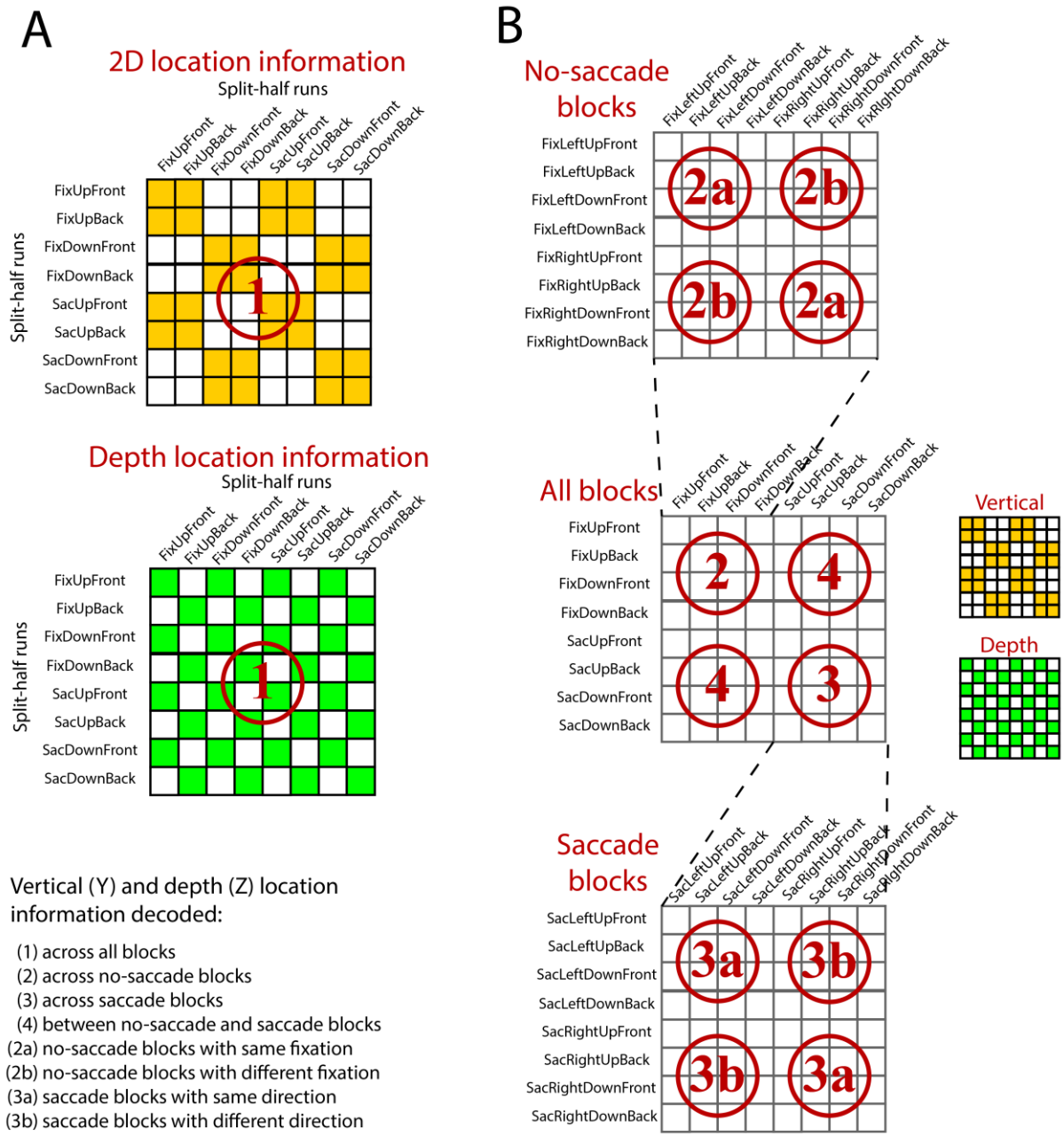

Figure 2 Hypothetical correlation matrices for MVPA. (A) Correlation matrices for calculating overall Y (top panel) and Z (bottom panel) location information, using data from all blocks. Information is calculated by subtracting between-category correlation coefficients (white cells) from within-category coefficients (colored cells). (B) For the subsequent breakdowns, Y and Z information are calculated in the same way (within-category minus between category), but only

from subsets of the cells in the correlation matrix. Each quadrant is numbered to link to the corresponding analyses, as indicated in the left bottom panel.

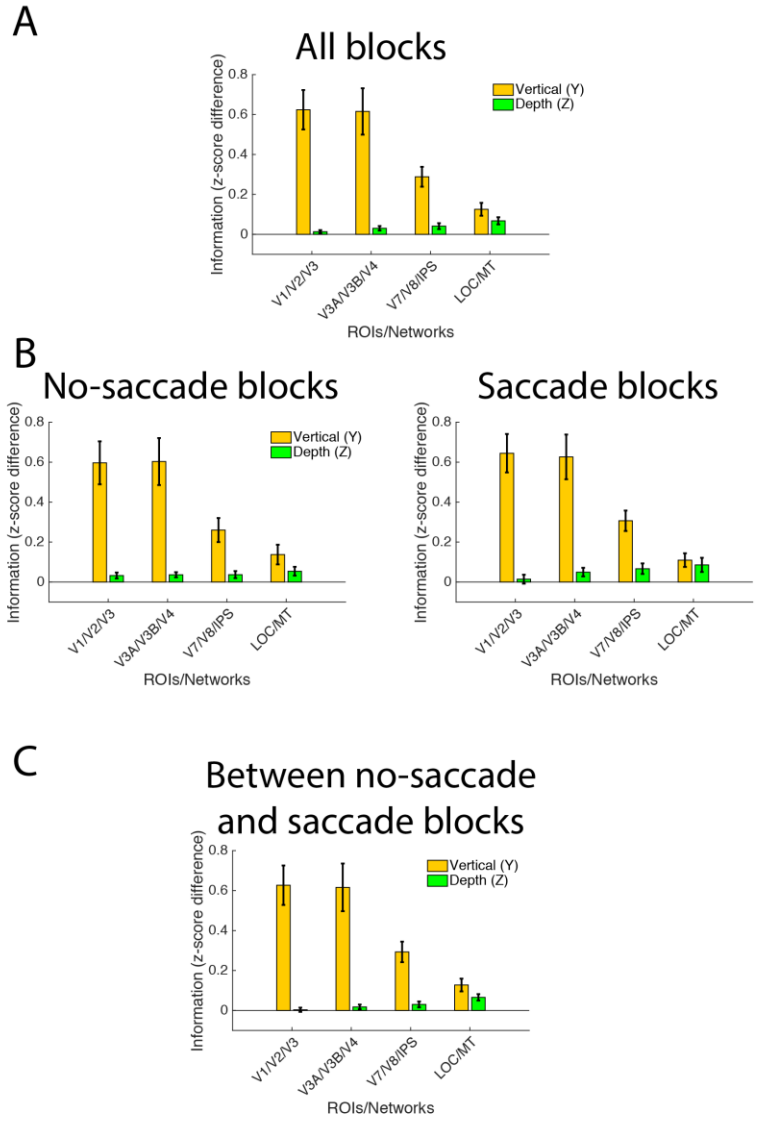

Figure 3 Y and Z location information averaged for the four grouped ROIs. (A) Information calculated with data from all blocks, using the correlation matrices in Figure 2 Matrix 1. (B) Information calculated with data from no-saccade blocks (left) and saccade blocks (right)

separately, using the correlation matrices in Figure 2 Matrix 2 and Matrix 3 respectively. (C) Information calculated with correlations between no-saccade and saccade blocks, using the correlation matrices in Figure 2 Matrix 4.

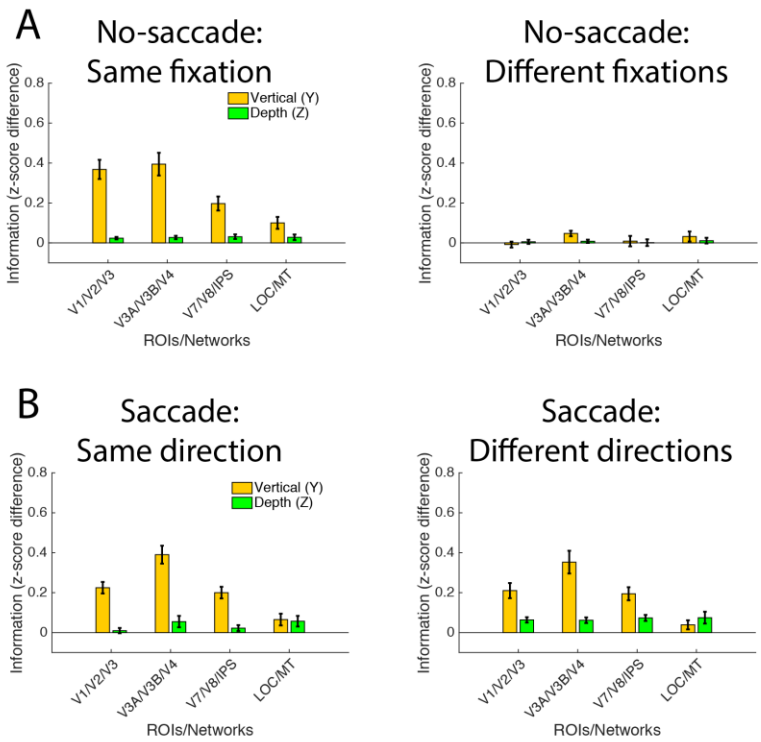

Figure 4 Tolerance/dependence analysis: Vertical (Y) and depth (Z) location information calculated separately for same vs different fixation position (no-saccade blocks) and saccade direction (saccade blocks). (A) No-saccade blocks. Y and Z location information calculated from MVPA between blocks with the same fixation position (left panel, using the matrix in Figure 2 Matrix 2a), and different fixation position (right panel, using the matrix in Figure 2 Matrix 2b). Y and Z location information is dependent on fixation position in no-saccade blocks. (B) Saccade blocks. Y and Z location information calculated from MVPA between sacade blocks with the same saccade directions (left panel, using the matrix in Figure 2 Matrix 3a), and

different saccade directions (right panel, using the matrix in Figure 2 Matrix 3b). Y and Z location information is tolerant of fixation differences in saccade blocks.

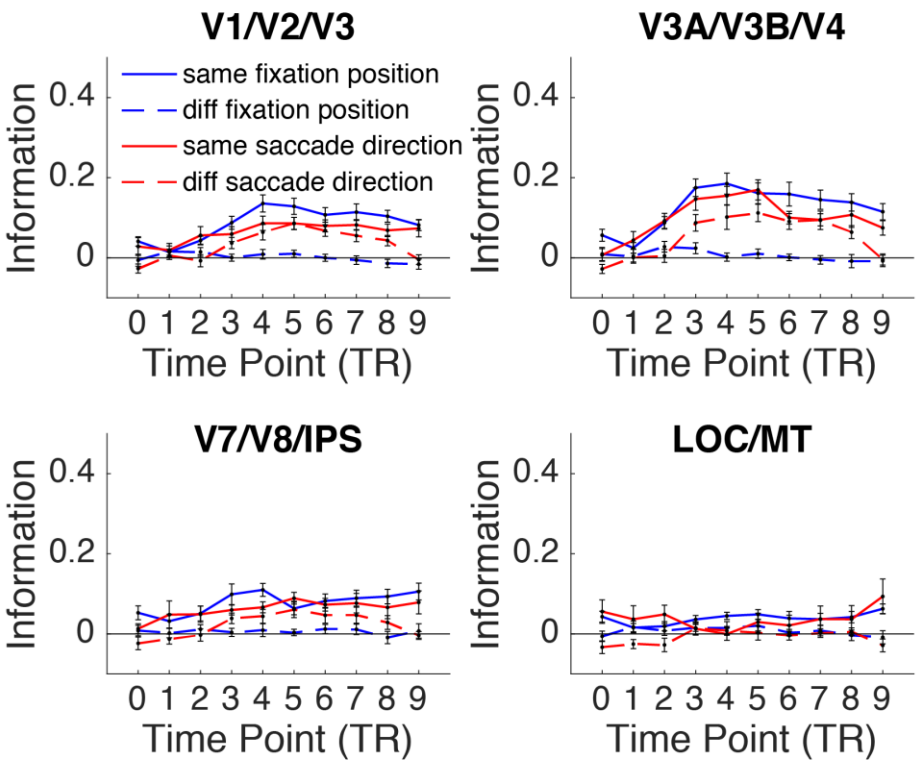

Figure 5 MVPA time course for vertical (Y) location information for each ROI group in the visual hierarchy (Z information timecourses shown in Supplemental Materials). Similar to Figure 5, in most regions location information is dependent on fixation position in no-saccade blocks but tolerant of fixation changes in saccade blocks, even when analyzing timepoint by timepoint.