Spatial tuning shifts increase the discriminability and fidelity of population codes in visual cortex

Running title: Spatial attention from units to populations

Vy A. Vo[1*], Thomas C. Sprague[1,2], John T. Serences[1,3,4*]
[1]Neurosciences Graduate Program, University of California, San Diego, La Jolla, CA 92093
[2]Department of Psychology and Center for Neural Science, New York University, New York, New York 10003
[3]Department of Psychology, University of California, San Diego, La Jolla, CA 92093
[4]Kavli Institute for Brain and Mind, University of California, San Diego, La Jolla, CA 92093

[*]Correspondence:
Neurosciences Graduate Program
University of California, San Diego
La Jolla, CA 92093-0634
vyaivo@ucsd.edu or jserences@ucsd.edu

Acknowledgements

Many thanks to the lab and particularly to Rosanne Rademaker and Edward Vul for comments on analyses and on the manuscript. This work was supported by National Science Foundation Graduate Research Fellowships to V.A.V. and T.C.S., a grant from the National Eye Institute (R01-EY025872) and a Scholar Award from the James S. McDonnell Foundation to J.T.S.

ABSTRACT

Selective visual attention enables organisms to enhance the representation of behaviorally relevant stimuli by altering the encoding properties of single receptive fields (RFs). Yet we know little about how the attentional modulations of single RFs contribute to the encoding of an entire visual scene. Addressing this issue requires (1) measuring a group of RFs that tile a continuous portion of visual space, (2) constructing a population-level measurement of spatial representations based on these RFs, and (3) linking how different types of RF attentional modulations change the population-level representation. To accomplish these aims, we used fMRI to characterize the responses of thousands of voxels in retinotopically organized human cortex. First, we found that the response modulations of voxel RFs (vRFs) depend on the spatial relationship between the RF center and the visual location of the attended target. Second, we used two analyses to assess the spatial encoding quality of a population of voxels. We found that attention increased fine spatial discriminability and representational fidelity near the attended target. Third, we linked these findings by manipulating the observed vRF attentional modulations and recomputing our population measures. Surprisingly, we discovered that attentional enhancements of population-level representations largely depend on position shifts of vRFs, rather than changes in size or gain. Our data suggest that position shifts of single RFs are a principal mechanism by which attention enhances population-level representations in visual cortex.

INTRODUCTION

Spatial receptive fields (RFs) are the basis of information processing throughout the visual system. They are directly modified by selective visual attention to improve the fidelity of sensory representations, which likely enables more precise and accurate behavioral choices (Desimone and Duncan, 1995; Anton-Erxleben and Carrasco, 2013). Prior macaque studies have found that covert spatial attention changes the position, size, and amplitude of responses from single-cell RFs in early cortical areas such as V1, V4, and MT by (Moran and Desimone, 1985; Connor et al., 1996, 1997, Womelsdorf et al., 2006, 2008; Roberts et al., 2007; David et al., 2008). Recent neuroimaging studies have also shown that single voxel RFs (vRFs) undergo similar response changes with attention, shifting towards the attended target or changing in size (Klein et al., 2014; Kay et al., 2015; Sheremata and Silver, 2015). Most accounts suggest that these RF changes improve the spatial representations of the attended target, either by boosting the signal-to-noise ratio (SNR) by increasing response amplitude, or by increasing the spatial resolution of the target representation by decreasing the size or tuning width (Desimone and Duncan, 1995; Anton-Erxleben and Carrasco, 2013; Cohen and Maunsell, 2014). These mechanisms are akin to turning up the volume knob (gain increase) or to using smaller pixels to encode a digital image (size decrease).

Despite all of these documented modulations, it is not yet clear how these different types of RF modulations are combined to support more robust population codes. Recent studies have only begun to explore how interactions between neurons may affect the coding properties of the population (Anton-Erxleben and Carrasco, 2013; Cohen and Maunsell, 2014). Yet analyzing these data at a population level is crucial for understanding how spatial attention changes the representation of an entire visual scene. Prior fMRI studies that measured many vRFs across space do not report the full pattern of response modulations across space, such as changes in the amplitude or size of vRFs with respect to the attended target (Sprague and Serences, 2013; Klein et al., 2014; Kay et al., 2015). The first aim of this study was to evaluate how all of these properties jointly change across space with spatial attention, in retinotopic regions from V1 through posterior IPS.

The second aim of the study was to evaluate how different types of RF modulations contribute to population-level enhancements of an attended region of space. Single RFs in early visual areas are fundamentally local encoding models that are relatively uninformative about spatial regions outside their immediate borders. To study their relationship to a population-level representation of space, we also need a larger-scale spatial encoding model which can incorporate information from many underlying, spatially selective RFs to form a representation of the entire visual scene. Here, we used two different population-level metrics of spatial encoding fidelity to investigate these questions. Specifically, we tested how changes in vRF amplitude, size, or position affect two measurements of population-level representations: (1) the spatial discriminability of population codes and (2) stimulus reconstructions based on a multivariate inverted encoding model.

We found that vRF position shifts increase both the spatial discriminability of the attended region as well as the fidelity of stimulus reconstructions near the attended target. Surprisingly, shifts in vRF position captured more of the population-level enhancements with attention than changes in vRF size or gain. This poses problems for traditional ‘labeled-line’ models of information processing, which posit that each RF in the visual pathway relies on consistently ‘labeled’ inputs from RFs in an earlier visual area. Our findings suggest that these shifts in the ‘labels’ of RFs play an important role in the attentional enhancement of visual information, and that labeled-line models may need to be reformulated to accommodate these data.

MATERIALS & METHODS

Task design and participants

We collected data from 9 human participants (4 female), 6 of whom had previously completed a set of retinotopic mapping scans in the lab (participants AA, AB, AC, AI, and AL in Sprague & Serences, 2013; participants AA, AC, and AI in (Sprague et al., 2014); all participants in (Ester et al., 2015). All participants provided written informed consent and were compensated for their time ($20/hour) as approved by the local UC San Diego Institutional Review Board. Participants practiced both the attention task and the localizer task before entering the scanner. A minimum of four hours of scanning was required to complete the entire analysis, so one participant was excluded due to insufficient data (they only completed 2 hours). Another participant was excluded for inconsistent behavioral performance, with average task accuracy at chance (48.6%). This yielded a total of 7 participants who completed the entire experiment (3 2-hour scan sessions per participant).

Participants centrally fixated a gray rectangular screen (120x90 cm) viewed via a head-coil mounted mirror (~3.85 m viewing distance). They attended one of three fixed locations on the screen: the fixation point or a target to the left or right. During each 2000 ms trial, subjects reported a change in the attention target. When subjects attended fixation, they reported whether a brief contrast change (100 – 400 ms, starting 300 – 1000 ms into the trial) was dimmer or brighter than the baseline contrast. The peripheral attention targets were two pentagons (0.17° radius; 50% contrast) centered 2.1° to the left and right of fixation (**Fig 1A**). When subjects attended a peripheral target, they reported whether it rotated clockwise or counter-clockwise (rotation duration 100 - 300 ms, starting 300 - 1600 ms into the trial). Inter trial intervals (ITIs) randomly varied between 1000 to 3000 ms in 500 ms increments (mean ITI: 2000 ms). The magnitude of the contrast change or the rotation was adjusted on each run to keep task performance for each participant near 75% (mean = 75.90%, bootstrapped 95% C.I. [72.46%, 79.20%]), with no significant difference between conditions as evaluated with a one-way repeated measures ANOVA randomization test (F(1,11) = 0.220, randomized p = 0.802). For four participants, we collected 6 runs on the attend periphery tasks without a change in the luminance of the fixation stimulus. Performance on the attend periphery tasks was stable across runs with and without the luminance change (repeated-measures ANOVA with run type x random participants factor; p = 0.439, null F distribution using randomized labels for 10,000 iterations). Therefore, these data were collapsed across scan sessions with and without changes in fixation luminance.

On 51 of the 61 trials in each run, a full-contrast 6 Hz flickering checkerboard (0.68° radius; 1.67 cycles/deg) appeared for 2000 ms at one of 51 different locations across the screen to map the spatial sensitivity of visually responsive voxels. When one of these checkerboards overlapped with any of the static attention targets, they were partially masked with a small circular aperture the same color as the screen background (0.16°/0.25° radius aperture for fixation/pentagon, respectively) that allowed the stimulus to remain seen. Participants were instructed to ignore the task-irrelevant flickering checkerboards throughout the experiment. During the 10 null trials, the participant continued to perform the attention task but no checkerboard was presented. Null trials and mapping stimulus trials were presented in a psuedorandom interleaved order.

The location of the checkerboard mapping stimulus on each trial was determined by generating an evenly spaced triangular grid (0.84° between grid points) and centering the checkerboard on one of these grid points. The location of the checkerboard mapping stimulus was then jittered a random amount from these grid points (+/- 0.42°/0.37° horizontal/vertical). When subjects attended the peripheral target, half of the runs were presented at the discrete grid positions so that we could achieve more stable stimulus reconstructions (see **Population analysis (2)**).

Magnetic resonance imaging

We obtained all structural and functional MR images from participants using a GE 3T MR750 scanner at University of California, San Diego. We collected all functional images (19.2 cm^2 FOV, 64 x 64 acquisition matrix, 35 interleaved slices, 3 mm^3 voxels with 0 mm slice gap, 128 volumes per scan run) using a gradient echo planar pulse sequence (2000 ms TR, 30 ms TE, 90° flip angle) and a 32-channel head coil (Nova Medical, Wilmington, MA). Five dummy scans preceded each functional run. A high-resolution structural image was acquired at the end of each session using a FSPGR T1-weighted pulse sequence (25.6 cm^2 FOV, 256 x 192 acquisition matrix, 8.136/3.172 ms TR/TE, 172 slices, 9° flip angle, 1 mm^3 voxels, 192 volumes). All functional scans were co-registered to the anatomical images acquired during the same session, and this anatomical was in turn co-registered to the anatomical acquired during the retinotopy scan.

EPI images were unwarped with a custom script from UCSD's Center for Functional Magnetic Resonance Imaging using FSL and AFNI. All subsequent preprocessing was performed in BrainVoyager 2.6.1, including slice-time correction, affine motion correction, and temporal high-pass filtering to remove slow signal drifts over the course of each session. Data were then transformed into Talairach space and resampled to 3x3x3 mm voxel size. Finally, the BOLD signal in each voxel was transformed into Z-scores on a scan-by-scan basis. All subsequent analyses were performed in MATLAB using custom scripts (to be available online upon acceptance).

We constrained our analyses to visually responsive voxels in occipital and parietal cortex using a separate localizer task (3-5 runs per participant). On 14 trials, participants fixated centrally and viewed a full-field flickering checkerboard (10 Hz, 11.0/8.3° width/height) for 8000 ms. Participants detected whether a small area (2D Gaussian, $\sigma = 0.2°$) within the checkerboard dimmed in contrast. Contrast dimming occurred between 500 to 4000 ms after the start of the trial, and lasted between 2000 to 3000 ms (all uniformly sampled in 500 ms steps). This contrast change occurred infrequently (randomly on 5 out of 14 trials) at a random location within the checkerboard. The average contrast change was varied between runs to maintain consistent performance at ~75% accuracy (mean performance 78.0%). On 8 trials participants simply fixated throughout the trial without a checkerboard being presented. For all subsequent analyses, only voxels in the retinotopically defined areas V1, V2, V3, V4, V3A/B and IPS0 with a significantly positive BOLD response to the localizer task (at FDR q = 0.05) were included (Benjamini and Yekutieli, 2001).

For all subsequent analyses, we used trial-wise BOLD z-scores. We estimated these by creating an event predictor for each checkerboard mapping stimulus and convolving it with a canonical two-gamma HRF (peak at 5 s, undershoot peak at 15 s, response undershoot ratio 6, response dispersion 1, undershoot dispersion 1). We then solved a general linear model (GLM) to find the response to each predictor. To standardize our data across runs, we z-scored the BOLD responses within each run and concatenated the z-scores across runs.

Statistical procedures

All reported confidence intervals (CIs) are computed by resampling the data with replacement (i.e. bootstrapping). The number of iterations for each bootstrapping procedure varied (depending on available computing power and time for that procedure) and are therefore reported with each result. For tests comparing a bootstrapped distribution against zero, p-values were computed by conducting two one-tailed tests against 0 (e.g., mean(param_change < 0) & mean(param_change > 0)) and doubling the smaller p-value. All repeated tests were FDR corrected (q = 0.05).

Voxel receptive field (vRF) estimation, fitting, and parameter analysis

We first estimated vRFs for each attention condition to investigate (1) how vRF parameters changed when participants attended to different locations and (2) the spatial pattern of vRF changes across visual space. We note here that prior reports have referred to similar voxel RF models as population receptive fields, or pRFs, to emphasize the fact that each voxel contains a population of spatially tuned neurons (Dumoulin and Wandell, 2008; Wandell and Winawer, 2015). However, since we are comparing modulations at different scales in the present study (i.e. modulations in single voxels and in patterns of responses across many voxels), we will refer to these single voxel measurements as voxel receptive fields (vRFs), and will reserve the term ‘population’ exclusively for multivariate measures involving several voxels, allowing our terminology to align with theories of population coding.

We estimated voxel receptive fields (vRFs) using a modified version of a previously described technique (Sprague and Serences, 2013). This method estimates a single voxel’s spatial sensitivity by modeling its BOLD responses as a linear combination of discrete, smooth spatial filters tiled evenly across the mapped portion of the visual field. These spatial filters (or spatial channels) form our modeled basis set. We then regressed the BOLD z-scores (*v* voxels x *n* trials) onto a design matrix with predicted channel responses for each trial (C, *k* channels x *n* trials) by solving Equation 1:

(1) $$B = WC$$

for the matrix W (*v* voxels x *k* channels).

Each of the *k* channels in the basis set was defined as a two-dimensional cosine that was fixed to reach 0 at a set distance from the filter center:

(2) $$f(r) = \left(0.5\left(\cos\left(\frac{r\pi}{s}\right) + 0.5\right)\right)^7 \ for\ r < s,$$

where *r* is the distance from the filter center and *s* is the size constant. Setting a zero baseline in this function ensured that we could estimate a stable baseline for each voxel by restricting the response of the channel to a known subregion of the visual display. Since the estimated vRF size depends on the size of the filters, we made the filters fairly small (1.08° FWHM) and dense (91 filters arranged in a 13 horizontal / 7 vertical grid, each spaced 0.83° apart). We then discretized the filters by sampling them in a high-resolution 2D grid of 135 by 101 pixels spanning 10° by 5°. The discretized filters (*k* filters by *p* pixels) were multiplied with a mask of the checkerboard stimulus on every trial (*p* pixels by *n* trials) so that the design matrix *C* contained predictions of the spatial channel responses on every trial of the mapping task.

In order to fit our estimated vRFs with a unimodal function, we used ridge regression to solve Equation 1. This is a common regularization method which sparsifies the regression solution by penalizing the regressors with many small weights (Lee et al., 2013). This meant solving for an estimate of *W* by the following:

(3) $$\widehat{W}^T = (CC^T + \lambda I)^{-1} CB^T,$$

where λ is the ridge parameter penalty term, and *I* is a *k* x *k* identity matrix. We estimated an optimal λ for each voxel by evaluating Equation 3 over a range of λ values (0 to 750) for all runs of the attention task (e.g., concatenating all attention conditions together). We then computed the Bayesian Information Criterion (BIC) for each of these λ values, estimating the degrees of freedom in the ridge regression as $df = trace(C\ (C^T C + \lambda I)^{-1} C^T)$. The λ with the smallest BIC was selected for each voxel. Since the attention comparisons are done within voxels, the varying λ penalty across voxels could not explain the attention effects we observed.

To select reliable voxels for analysis, we next implemented a set of conservative thresholding steps (**Table 1**). We first needed to select voxels with reliable visual responses, so we only kept voxels with trial beta weights that predicted at least 50% of the BOLD time courses in each scan session. Second, we only used voxels that could be successfully regularized with ridge regression. Any voxels with the maximum λ (750) were discarded, as this indicated that the ridge regression solution had not converged. Finally, we verified that the resulting regression model could predict an independent dataset, so we performed exhaustive leave-one-run-out cross validation for each attention condition. This ensured that the λ estimated across attention conditions produced reliable data for each condition separately. We estimated *W* using data from all but one run (Equation 3) and used this to predict the BOLD GLM trial estimate of the left-out run (Equation 2), again all done separately for each condition. We then computed the mean correlation between the predicted & real BOLD GLM trial estimates across cross-validation folds for each voxel. Note that it is not possible to calculate a coefficient of determination on regularized data, since the process of ridge regression changes the scale of the predicted data (see Huth et al., 2012 for more). We only kept voxels where this cross-validation r > 0.25 for all 3 conditions.

To quantify each vRF, we fit each voxel with a smooth 2D function with 4 parameters: center, size, baseline, and amplitude (**Fig 1b**; Equation 2). Here, we define the vRF baseline as the voxel's typical response that does not depend on the position of the mapping stimulus. The vRF amplitude is defined as the spatially-selective increase in a voxel's response above this baseline. Together, these two parameters index how much of the voxel's response is due to a change in mapping stimulus position. Finally, the size and location parameters estimate the spatial selectivity and the spatial position preference of the vRFs, respectively. We first downsampled the vRFs by multiplying the estimated weights $\widehat{W}$ for each voxel (a 1 x *k* channel vector) by a smaller version of the spatial grid that contained the basis set (68 by 51 pixel grid; 10° by 5°). This speeded up the process of fitting the pixelwise surface with Eq. 2. This fitting process began with a coarse grid search that first found the best fit in a discrete grid of possible vRF parameters (center sampled in 1° steps over the mapped portion of the visual field; size constant logarithmically sampled at 20 points between 2.3 and 38.5, which gives FWHMs between 0.9° and 15.3°). We then estimated the best fit amplitude and baseline for each of the grid points using linear regression. The grid point fit with the smallest root mean square error (RMSE) provided the initialization seed to a continuous error function optimization algorithm (fmincon in MATLAB). This fit had several constraints: the final solution must place the center within 2 grid points of the seeded fit (parameterized by position and size) and within the mapped visual field; the amplitude must be between 0 and 5; the baseline must be between -5 and 5 BOLD z-score units. Occasionally, this nonlinear fitting algorithm did not converge and resulted in a larger error than the original seed. In this case we took the best fit grid point as the final fit.

To test whether vRF fit parameters changed when participants focused spatial attention at different positions, we compared fits during each attend periphery condition with fits during the attend fixation condition. We computed a difference score (attend peripheral – attend fixation) to describe the magnitude of the attentional modulation. For example, a difference score of –2° in the FWHM of the vRF would indicate that the response profile width decreased when the participant attended to the peripheral target. We then tested whether the vRF parameter difference scores differed significantly from 0 within a visual region of interest (ROI) by bootstrapping the distribution of difference scores across participants 10,000 times.

To test whether these vRF changes were modulated by their position in the visual field, we first calculated each vRF's distance from the attended location (*v_dist_attn*) using its position during the fixation task. These were sorted into distance bins (0° to 2.5°, in 0.25° steps) and all vRF difference scores in that bin were averaged across participants. We then fit an *n*th order polynomial to the binned vRF difference scores as a function of *v_dist_attn*, where $n = 0, 1$, or 2. This corresponds to a constant offset (0^{th} order), a linear fit (1^{st} order), or a quadratic or parabolic fit (2^{nd} order). The most parsimonious fit was chosen using a nested F-test. Fit coefficient CIs were generated by bootstrapping the data across participants 5,000 times before repeating the binning, averaging, and fitting procedure.

Population analysis (1): Fine spatial discriminability metric

To compute the spatial discriminability of a population of vRFs, we estimated the spatial derivative of each vRF at every point in the mapped visual field in 0.1° steps (**Fig 1C**). This was done by taking the slope of the vRF along the x and y direction at each pixel in the image of the visual field and squaring this value (Scolari and Serences, 2009, 2010). This measurement is a good descriptor of how well a population code can discriminate small changes in the spatial arrangement of the stimulus array, which depends on the rising and falling edges of a tuning curve rather than the difference between the peak response and a baseline response (Regan and Beverley, 1985; Pouget et al., 2003; Butts and Goldman, 2006; Navalpakkam and Itti, 2007; Scolari and Serences, 2009, 2010). In order to restrict our measurements to the relevant area near the peripheral target, we computed discriminability values within 1 degree of the center of each target across both spatial dimensions (x and y). These were summed and divided by the maximum discriminability value in that population in order to make the results comparable despite changes in vRF coverage or responsiveness.

Population measurements (2): Stimulus reconstructions using an inverted spatial encoding model

In addition to computing the discriminability metric described above, we also reconstructed an image of the entire visual field on each trial using a population-level encoding model. Compared to the local spatial discriminability index, this is a more sensitive method of assessing the amount of spatial information encoded in a population of voxels because it exploits the pattern of response differences across voxels, rather than treating each voxel as an independent encoding unit (Serences and Saproo, 2012; Sprague et al., 2015).

We train the spatial encoding model using a procedure similar to the vRF estimation analysis described above (**Fig 4a**). This yields an estimated matrix of weights, $\widehat{W_2}$, which specifies how much each voxel in a region of interest responds to each of the spatial channels (Brouwer and Heeger, 2009; Serences and Saproo, 2012; Sprague and Serences, 2013; Sprague et al., 2015). We then solve Eq. 1 using the Moore-Penrose pseudoinverse with no regularization:

$$(4) \qquad \widehat{W_2} = BC^T(CC^T)^{-1}$$

C was constructed using a set of 48 evenly tiled spatial filters (Eq. 2; 8 horizontal / 6 vertical; spaced 1.43° apart; 1.78° FWHM). $\widehat{W_2}$ was estimated using the data from the jittered position runs. This was done separately for each participant, using a balanced training set that contained an equal number of attend left, attend right, and attend fixation runs.

To reconstruct a representation of the mapped visual space, we inverted the model and multiplied the pseudoinverse of the estimated weight matrix $\widehat{W_2}$ with a test dataset from the discrete position runs (B_2), yielding estimated channel activations for each trial (C_2; k_2 channels by *t* test trials) (Equation 5). Thus, we refer to this analysis as the inverted encoding model (IEM).

(5) $$\hat{C}_2 = \left(\widehat{W_2}^T \widehat{W_2}\right)^{-1} \widehat{W_2}^T B_2$$

Because of mathematical constraints on inverting W_2 (number of voxels must be greater than number of channels), we included all voxels in each ROI instead of just the subset of well-fit voxels used in the vRF analyses described above. We performed this inverting procedure twice using different test datasets, once for the discrete position attend left runs and once for the discrete position attend right runs.

When we multiply the resulting channel activations by a grid of pixels that define the spatial channels, we obtain a spatial representation of the entire visual field on each trial. This image contains a stimulus reconstruction showing where the checkerboard should have been given the trained model and the activation pattern across all voxels in the independent test set. The stimulus reconstructions were then fit in the same manner as the vRFs, using Eq. 1 to estimate the center, size, amplitude, and baseline (mean fit RMSE across all ROI reconstructions 0.173; 95% CI [0.102, 0.312]). Here, the baseline is an estimate of the multivariate reconstruction that is spatially non-selective—i.e., not significantly modulated by the position of the mapping stimulus. The amplitude describes the maximal increase in that reconstruction relative to baseline when the mapping stimulus is on the screen.

To assess how attention changed reconstructions of the mapping stimulus across the visual field, we first computed a difference score that described the effect of attention by folding the visual field in half (i.e. collapsing across hemifield) and comparing parameters in the attended vs. ignored hemifield. We excluded the reconstructions that fell along the vertical meridian (3 of 51 stimulus positions). This allowed us to control for the overall effect of eccentricity while remaining sensitive to other spatial patterns in stimulus reconstruction modulations.

We then set up a single factor repeated measures omnibus ANOVA to determine which pairs of ROI and parameter (e.g., V1 size, V1 amplitude, etc.) were affected by either attention or Euclidean distance from the target stimuli. The attention factor had two levels (attend/ignore) and the distance factor had 6 levels (6 evenly spaced distance bins from 0° to 2.54°). Based on the results of this omnibus test, we tested any significant ROI-parameter combination in a 2-way repeated measures ANOVA of attention by distance. To estimate the p-values for these tests, we generated empirical null distributions of the F-scores by randomizing the labels within each factor 10,000 times within each participant. Reported p-values are the percentage of the randomized F-scores that are greater than or equal to the real F-scores.

Population analysis (3): Layered spatial encoding model to link vRFs to multivariate stimulus reconstructions

In order to test how changes in the response properties of the underlying vRFs contributed to changes in the fidelity of region-level stimulus reconstructions, we generated simulated patterns of voxel activity on every trial by predicting the response to each stimulus based on the vRF fit parameters. We then used this simulated data to estimate and invert a population-level spatial encoding model, as described above (**Fig 6A**). Note that for these simulations, we could only use well-fit voxels to generate simulated BOLD timeseries. Therefore, we could not accurately estimate reconstructions for some participant-ROI pairs that had an insufficient number of voxels. Pairs that indicated a poorly conditioned matrix (e.g., number of voxels is fewer than the number of channels) were excluded (total 10 out of 35 pairs; V3 (AI, AP); V3A/B (AL, AP, AT); V4 (AA, AT); IPS0 (AA, AR, AU)).

To simulate each voxel's BOLD response on every trial that the participant completed in the real experiment, we first created a high-resolution set of spatial channels (21 by 11 channels spaced 0.5° apart, FWHM = 0.65°) and generated weights for each channel based on the vRF fit obtained from prior analysis. That is, we evaluated Eq. 2 for each channel at the vRF's fit center and adjusted the response gain by multiplying this result by the fit amplitude and adding the fit baseline. We then added independent Gaussian noise to each of these channel weights, simulating a small amount of variance in the voxel's response ($\sigma = 0.5$). Each voxel's channel weights were then multiplied by the idealized channel response on each trial (that is, the channel filter convolved with the stimulus mask), effectively simulating the BOLD response on each trial for the entire population of voxels according to their measured vRFs. We added Gaussian noise to this simulated response as well ($\sigma = 0.5$). We then computed stimulus reconstructions using the same method as described above (the IEM in **Population analysis (2)**), averaging resulting reconstructions across participants and like positions before fitting.

To assess the stability of the reconstructions that were based on simulated data, we repeated the simulations 100 times and averaged across the fits of all iterations to generate the plots in **Fig 6**. Note also that the chosen level of noise did not qualitatively impact the results. For example, rather than just adding Gaussian noise, we also created a noise model that followed the covariance structure between all voxels in each ROI. To estimate the covariance matrix, we computed the residuals between the true trial-wise beta weights and the predicted trial-wise beta weights for each voxel based on its vRF model. We then calculated the pairwise covariance between the residuals for each set of voxels. Last, we added noise that followed this covariance structure to each voxel's channel weights and simulated BOLD response. This noise was scaled to be the same as the noise level that most accurately captured the real reconstruction data (i.e., mean noise is 0.5 standard units). The pattern of results between each of the model manipulations remained the same, so those results are not discussed here.

To compare whether the results of the layered model differed significantly from the reconstructions generated with real data, we first calculated difference scores across attention condition (attended – ignored; see **Population analysis (2)**). This yielded 24 difference scores each for both attention conditions. Since the real data did not have any repeated iterations, we averaged across all 100 iterations of the model to match the dimensionality of the real reconstructions (2 conditions x 24 difference scores x 4 parameters). We then calculated the error between the difference scores from the full empirical dataset (as the data shown in **Fig 5**) and the modeled data to obtain the root mean square error (RMSE).

To test how shifts in vRF centers contributed to changes in the stimulus reconstructions, we also generated reconstructions from modeled voxels that had the same fit center across both attention conditions. We defined each voxel's vRF center as the fit center obtained from that voxel during the neutral attend fixation condition. We then repeated the stimulus reconstruction, reconstruction fitting, and statistical testing as described above. A similar procedure was repeated for all reported combinations of parameter changes across conditions. Again, whichever parameter was held constant took its value from the neutral attend fixation condition.

To calculate the confidence intervals on the RMSE changes in **Fig 6C**, we resampled with replacement across the 100 model iterations and took the difference between the RMSE of the null model, in which no parameters varied between attention conditions, and the RMSE of the model which held some number of vRF parameters constant across attention conditions. This procedure was repeated 500 times.

RESULTS

Modulations of vRF properties with spatial attention

We estimated single voxel receptive fields (vRFs) for each voxel in 6 retinotopically-identified visual areas from V1 to IPS0. The estimation of vRFs was done independently for each attention condition so that we could compare a single voxel's spatial tuning across conditions.

To confirm that the fit sizes were consistent with previous results, we fit a line to the estimated sizes as a function of the vRF center eccentricity. First, we combined all vRFs across participants and conditions in each ROI. We then binned the vRF centers every 0.25° from fixation and calculated the mean size (**Fig 2b**). We first replicated an increase in vRF size with increasing eccentricity, and an increase in the slope of this relationship across visual regions (Gattass et al., 2005; Dumoulin and Wandell, 2008; Amano et al., 2009; Harvey and Dumoulin, 2011) (**Fig 2b**). These observations confirm that our method produced reasonable vRF estimates that were consistent with previous reports.

Covert attention to either the left or the right position modulated vRF properties by shifting vRFs significantly closer to the attended location compared to the attend fixation condition ($p < 0.005$ in all ROIs). While we did observe size changes in individual voxels, the mean change was not significantly different from zero ($p > 0.05$ in all ROIs). Size increases have been previously reported in tasks that required subjects to attend to the mapping stimulus rather than to ignore it, as in the present study (Sprague and Serences, 2013; Kay et al., 2015; Sheremata and Silver, 2015). In these previous studies, the locus of attention changes on each trial. Accordingly, if attention attracts RFs as our data suggest, these combined shifts might average out to form a larger RF estimate. Here, we fixed the locus of attention so we could more finely characterize the effects of focal attention and found no net change in vRF size. However, we did find a general increase in vRF response gain, such that amplitude increased while mean baseline decreased ($p < 0.001$ for all tests). Since all of these measures were calculated relative to a fixation task, these data suggest that covert spatial attention to a peripheral location caused widespread position and gain modulations in all vRFs across the visual field.

Previous reports in humans and monkeys have suggested that the preferred position of RFs shift when subjects covertly attend to an area in the visual field (Womelsdorf et al., 2006, 2008; Klein et al., 2014) and when they make visually-guided saccades to an attended location (Zirnsak et al., 2014). It is unclear, however, whether these position shifts all radially converge towards the attended target or whether the RFs shift uniformly along a vector extending from fixation to the attention or saccade target (Tolias et al., 2001; Klein et al., 2014; Zirnsak et al., 2014). Furthermore, reports of other RF properties (such as size) modulating with attention have been mixed (Connor et al., 1996, 1997; Womelsdorf et al., 2008; Niebergall et al., 2011; Sprague and Serences, 2013; Klein et al., 2014; Kay et al., 2015; Sheremata and Silver, 2015). We therefore examined whether each of the vRF parameter changes was dependent on the vRF’s location in the visual field, relative to the attended location. First, we created radial distance bins centered on the left or right attended locations, and sorted voxels into these bins based on their preferred position during the fixation condition. After this sorting procedure, data from the right condition were flipped and collapsed with the left condition.

When we plotted vRF position changes in each bin, we found that spatial attention caused vRF position shifts that converged on the attended location (two-tailed sign test on vector direction, $p < .001$ in all ROIs). That is, vRFs shifted closer to the attended location (**Fig 2c**), compared to when subjects attended fixation (mean shift across all vRFs and ROIs: -0.244, bootstrapped 95% C.I. [-0.607, -0.038], **Fig 2d**). Note that small eye movements toward the attended location cannot explain receptive field convergence: this would cause all vRFs to shift in the same horizontal direction, rather than radially converging on one point. These data are consistent with results from both humans (Klein et al., 2014) and macaques (Connor et al., 1996, 1997, Womelsdorf et al., 2006, 2008). However, the prior study in humans focused only on vRFs with preferred locations that were foveal to the attended location, and the studies in macaques only report RF position changes in V4 and MT. By contrast, we show that vRF centers converge on the attended location across all visual areas, including primary visual cortex, as well as in vRFs with centers peripheral to the attended location.

These plots (**Fig 2a, 2d**) also suggested that vRFs farther from the attended location underwent larger position changes with covert shifts of spatial attention. The size of the attentional modulation may be dependent on the distance between the vRF center and the attended target. To test for this, we fit a polynomial to the vRF parameter changes as a function of distance from the attended location (**Materials and Methods**). We selected the most parsimonious fit ranging from a mean change in vRF parameter (0^{th} order polynomial) to a parabolic change (2nd order polynomial) by conducting a nested F-test (**Table 2**). The significant polynomials of order n > 0 are plotted in **Fig 2e**.

Position changes with attention were significantly modulated by the initial vRF position (relative to the locus of attention) in higher visual areas such as V4, as indicated by a significant slope coefficient from a linear fit. While all size changes were best described by a quadratic function, only area V2 showed a significant modulation: we observed size increases in V2 vRFs near the attended location, size decreases about 1 degree away from the attended location, and size increases again for vRFs farther away. Amplitude changes with attention were significant in visual areas higher up in the visual hierarchy, namely V3A/B and IPS0. In these visual areas, amplitude increased for vRFs farther from the attended location. This was paired with a large mean decrease in baseline (**Fig 2d**). These tests suggest that the spatial relationship between the vRF and the attended target changes the type and magnitude of the attentional modulation in different visual areas, consistent with findings from macaque neurophysiology (Connor et al., 1996; Niebergall et al., 2011).

Increases in spatial discriminability depend primarily on vRF position shifts

Next, we assessed how different types of RF modulations influenced the precision of population-level codes for spatial position. We first computed a discriminability metric that described the ability of a population of tuning curves to make fine spatial judgments (**Materials and Methods**). Spatial discriminability near the attended target increased relative to the ignored target in the opposite visual hemifield (**Fig 3a**).

We then tested how different types of vRF modulations (such as size changes or position shifts) affected this spatial discriminability metric. As a baseline comparison, we first computed discriminability based on vRFs estimated during the attend fixation runs for each subject. We then added different sets of observed attentional modulations to the population before recomputing spatial discriminability. For example, we shifted all the vRF centers to match the measurements when a subject was attending to the left target and computed discriminability near the attended target. Since the response baseline of a vRF does not affect the discriminability metric, we excluded this type of attentional modulation from these analyses.

Across all ROIs, we found that vRF position shifts played the biggest role in increasing fine spatial discriminability compared to changes in size or changes in amplitude (**Fig 3b**). Position modulations alone led to a large increase in spatial discriminability, while other combinations of parameter modulations only had an impact if we added in position shifts (i.e. a change in size and position increased discriminability, but size alone did not). The only departure from these patterns was observed in IPS0, where a combination of amplitude and size produced an increase in discriminability even in the absence of changes in vRF position.

Spatial attention increases the fidelity of population-level stimulus reconstructions

By design, the spatial discriminability metric we computed is only informative about local spatial representations, and cannot assess how different patterns of vRF modulations might result in representational changes across the visual field. To address this point, we built a multivariate spatial encoding model to measure how attention changes the representations of visual information in disparate parts of space. This also allowed us to further test the effects of vRF modulations on the encoding properties of the population, including response baseline changes that were not captured by our discriminability metric.

The spatial inverted encoding model (IEM) reconstructed an image of the entire visual field on each test trial. We first trained the model using the responses of each voxel on a set of training trials with known mapping stimulus positions. We then created image reconstructions on independent test trials by inverting the model and multiplying it by the voxel responses during each test trial (**Fig 4a**; **Materials and Methods**). Each image contained a representation of where the mapping stimulus should have been given the pattern of voxel activations on that particular trial. The IEM successfully reconstructed the task-irrelevant mapping stimuli using activation patterns across voxels in each visual area from V1 through IPS0 (**Fig 4b**; grand mean error between fit and actual position 2.36°, 95% CI [0.56°, 4.89]).

We used these stimulus reconstructions as a proxy for the quality of the spatial representations encoded in a population of voxels. This is line with previous studies showing that stimulus reconstructions have change in amplitude or size as a function of cognitive demands. (Brouwer and Heeger, 2013; Ester et al., 2013; Sprague and Serences, 2013; Sprague et al., 2014, 2015, 2016).

First, we compared how reconstructed representations of each mapping stimulus changed as subjects shifted their spatial attention. We ran a repeated measures ANOVA of *attention* x *distance bin* for each reconstruction fit parameter (**Materials and Methods**). Here, a main effect of attention would suggest that stimulus reconstructions in the attended hemifield changed in a consistent way compared to the ignored hemifield. A main effect of distance would suggest that stimulus reconstruction changes had a consistent spatial pattern across both the attended and ignored hemifields. This would occur when a stimulus' representation was altered with distance from the attention target. For example, the stimulus reconstruction center should vary linearly with the stimulus' true distance from the attention target. And lastly, an interaction effect would suggest that the distance effect was dependent on whether the reconstruction belonged to the attended or ignored hemifield. In our task, the reconstructed stimuli are always irrelevant to the task of the observer. We therefore predicted an interaction effect where spatial attention would selectively modulate stimulus reconstructions around the attended location (Connor et al., 1996, 1997).

We found that reconstruction amplitude was selectively increased near the attended location in V3, V4, V3A/B, and IPS0 (interaction effect, bootstrapped p < 0.005; **Fig 5**; **Table 3**). This can be interpreted as a local boost in SNR. Prior reports found that attending to the mapping stimulus – as opposed to attending to a peripheral target as in the current experiment – caused an increase in the amplitude of all stimulus reconstructions (Sprague and Serences, 2013). That is, representations of task-relevant stimuli increased in SNR. We find here that even representations of task-*irrelevant* stimuli near the attended region of space increase in amplitude, consistent with the idea of an attentional 'spotlight' which boosts the fidelity of spatial representations near the attention target.

Although the amplitude interaction effect was present in most visual areas we tested (**Fig 5**), we found other effects limited to V3A/B and IPS0 that involved modulations in stimulus representations in the ignored hemifield. In these regions, we found that stimulus reconstructions in the ignored hemifield shifted away from the ignored target location (interaction, bootstrapped p < 0.005). We also observed a relative size increase near the ignored attention stimulus in IPS0 (interaction, bootstrapped p = 0.005). These results suggest that stimulus reconstructions in the ignored hemifield are less spatially precise in posterior parietal cortex. Finally, there was also a main effect of attention on reconstruction size and baseline in areas V4 & V3A/B (bootstrapped p's <= 0.005). However, unlike the interaction effect in IPS0, size changes in V4 and V3A/B did not vary as a function of distance between the reconstruction and the attended target location.

Using a layered encoding model to explore how single voxel RFs change population-level codes

In our final analysis, we used a layered spatial encoding model to determine how changes in vRF properties affected the representations of mapping stimuli in the multivariate reconstructions discussed in the previous section (**Fig 1c**; **Fig 4a**). The goal of this analysis was to determine which vRF modulations contribute the most to the observed increase in the amplitude of stimulus representations around the attended location (**Fig 5**). This analysis thus complements our analysis of the spatial discriminability metric which demonstrated that vRF position changes significantly increased the ability of the population to make fine spatial discriminations near the attention target (**Fig 3c**).

The layered spatial encoding model we built links the response properties of single voxels to the encoding properties of a whole population of voxels in a region of visual cortex (**Fig 6a**). In the first layer of the model, we used the fit vRFs to generate simulated BOLD data from each voxel under different attention conditions. We then repeated the multivoxel stimulus reconstruction analysis on this simulated data to model population results for the second layer of the model. This approach allowed us to perform virtual experiments to test how changes in the first layer impacted the second layer. That is, we manipulated which vRF parameters changed with attention (first layer) and observed the resulting changes in the population-based stimulus reconstructions (second layer). For example, we could test whether an overall increase in vRF response gain with attention would be necessary or sufficient to reproduce the amplitude increases observed in the empirical stimulus reconstructions reported above. These virtual experiments also allowed us to compare the relative impact of one type of response modulation (e.g. size changes) with other types of response modulations (e.g. position shifts).

We first validated our layered model to ensure that it produced results that matched the empirical data. Since our procedure only allowed us to use voxels with reliable vRF fits, we compared the results of the layered model to stimulus reconstructions in the full empirical dataset reported above. **Table 4** reports the RMSE between the fits to the layered IEM reconstructions and the full empirical IEM reconstructions. Our model reproduced the main pattern of results we observed in the previous section. In particular, covert attention led to an increase in the amplitude of reconstructions near the locus of attention, suggesting that our simulated data sets accurately captured the main modulation observed in the real data (**Fig 6b**, gray bars).

We then compared two basic manipulations of the layered IEM to see how they contributed to the amplitude increase that we observed in the stimulus reconstructions. When we abolished the position shift between attention conditions in the first layer of the model, we observed a decrease in stimulus reconstruction amplitude (**Fig 6b**). This suggests that spatial position shifts at the single voxel level are necessary for amplitude changes in stimulus reconstructions. When we held vRF sizes constant across attention conditions, there was little change in the amplitude effect in most ROIs, suggesting that size changes in vRFs were not necessary for changes in stimulus reconstruction amplitude.

To more formally compare each manipulation of the layered IEM, we compared each model to a baseline in which no vRFs changed with attention (far left in **Fig 6c**). This baseline should have the highest RMSE, and any additional attentional modulations to the underlying vRFs should decrease the error between the simulated data and the empirical data. Conversely, a model with higher RMSE is worse at accounting for the empirical data. In all ROIs, a model that abolished position shifts had a higher RMSE than a model which abolished size shifts (**Fig 6c**, light red and green bars). In fact, just modeling vRF position shifts was sufficient to significantly decrease RMSE in all ROIs except V4. However, this is likely because the layered IEM was a poor model for the attention effects in V4. This is evidenced by the fact that the baseline model did not have the highest RMSE (**Fig 6c**).

The overall pattern of results across ROIs is consistent with the interpretation that shifts in the position of vRFs have the largest impact on the population-level representations, while changes in vRF size or gain play smaller roles in changing the fidelity of the population code.

DISCUSSION (Max 1500 words; current 1437)

By simultaneously measuring the response properties of both single voxels and populations of voxels within retinotopic areas of visual cortex, we were able to link attentional modulations of spatial encoding properties across scales. Our data provide an initial account of how different types of RF modulations improve the quality of spatial population codes. We first report how different types of vRF modulations depended on the distance between the vRF's preferred position and the static attention target (**Fig 2**). We then found that shifts in the preferred position of vRFs near the attended target increased the spatial discriminability of a population (**Fig 3**), as well as the amplitude of stimulus reconstructions based on populations of vRF responses (**Fig 5**).

Attentional modulations of spatial RFs

While our study is not the first to simultaneously measure a population of RFs tiling a continuous portion of visual space, we provide new data on how vRF responses are modulated around a covertly attended static target (Sprague and Serences, 2013; Klein et al., 2014; Kay et al., 2015; Sheremata and Silver, 2015). Like prior macaque studies, we find that the spatial pattern of attentional modulations is widely variable, but that position shifts in RFs depend heavily on the distance from the attended target (Connor et al., 1996, 1997). We also found that vRF size modulations weakly depended on the RF's spatial relationship to the attended target, even though population-averaged size changes remained at a constant level across voxels (**Fig 2d-e**). Comparison to the existing literature suggests that patterns of RF size modulations likely depend on the nature of the spatial attention task. In fMRI tasks where subjects attended to the mapping stimulus itself, researchers report average vRF size increases with attention (Sprague and Serences, 2013; Kay et al., 2015; Sheremata and Silver, 2015). Furthermore, the relative size of the attention target and mapping stimulus likely play a key role as well. In macaques, RFs in area MT shrink when measured with a mapping probe smaller than the stimulus, but increase in size when macaques track the mapping probes as they move across the screen (Womelsdorf et al., 2006, 2008; Anton-Erxleben et al., 2009; Niebergall et al., 2011). Taken together, these observations demonstrate that the pattern of response modulations in single cells and in single voxels depends on the spatial relationship between the attended target and the spatial extent of the encoding unit.

We note that while the similarity between attentional modulations of single cell RFs and single voxel RFs is compelling, they are not interchangeable. Given that fMRI voxels in retinotopically organized regions of visual cortex sample from a broad array of neurons with roughly the same spatial tuning preferences, a position shift in a vRF could be driven by either a change in the preferred position of single neurons, or by a change in the gain profile across neurons tuned to slightly different locations in the visual field. This principle also holds true of position shifts in single neuron RFs, since cortical neurons typically receive input from those with smaller RFs in earlier visual areas (McAdams and Maunsell, 1999; Baruch and Yeshurun, 2014; Dhruv and Carandini, 2014). However, our population-level analyses do demonstrate how the properties of local encoding units–both single-cell and single-voxel RFs–might contribute to population-level representations of space.

Attention boosts the spatial encoding fidelity of a population

We first measured the overall capacity of a population of voxels to make fine spatial discriminations in a region of space. We found that attention increased spatial discriminability near the attended target, relative to the ignored target. We then performed virtual experiments on the underlying vRFs contributing to the population to determine how they affected the spatial discriminability metric. We report that vRF position shifts increased spatial discriminability significantly more than vRF size changes or even gain changes (**Fig 3**). As above, we note that spatially-specific patterns of gain changes in input RFs could produce these position shifts in downstream neural populations. This observation suggests that gain modulations with attention may exert their largest effects on the downstream population, where these patterns of gain changes become apparent shifts in vRF position. Our data are consistent with the interpretation that a neural population only begins to encode the attended area with higher fidelity after input gain changes are transformed into apparent shifts in spatial tuning in the encoding units of that population.

Since the spatial discriminability metric is only informative about a local portion of space, we performed a second population analysis to reconstruct an image of the entire visual field on each trial using a multivariate IEM. Attention increased the amplitude of stimulus reconstructions near the attention target, indicating an increase in representational fidelity that accompanies the change in spatial discriminability. In addition, the layered spatial encoding model revealed that shifts in vRF position could sufficiently account for these attentional enhancements in the population-level stimulus reconstructions, but changes in vRF size could not. Altogether, our data demonstrate that spatial tuning shifts in a group of RFs may be the dominant way that single encoding units alter the properties of a population spatial code.

Our findings also underscore the fact that changes in the spatial encoding properties of single units do not directly translate into analogous changes in the encoding properties of a population of those same units. This is particularly true when considering the effects of spatial attention on representations of the entire visual scene. Although we found that single units shifted their preferred position towards the attended target, population-level representations did not generally shift with attention. When the population code did shift its encoded position for a given stimulus, we found that it was typically representations of the ignored stimulus that shifted farther from the true stimulus location (**Fig 5**), consistent with a tendency towards more error-prone representations of stimuli far from the relevant portion of the screen. These types of differences further emphasize the need to understand the effects of cognitive state and task demands on population codes for the entire visual scene, rather than focusing solely on single units encoding largely local visual information.

Tuning shifts and labeled lines

Historically, shifts in the tuning of a RF have not been considered one of the main mechanisms by which attention modulates population-level information, although a handful of recent papers suggest that this view is being reconsidered (David et al., 2008; Anton-Erxleben and Carrasco, 2013). This is largely due to ‘labeled-line’ theories of visual information processing, which posits that a single neuron has a consistent feature label which downstream neurons rely on to perform computations and transmit stable information (Barlow, 1972; Doetsch, 2000; David et al., 2008). When a spatial RF shifts position as a function of cognitive state (e.g., attention), that single neuron’s feature label is no longer consistent. Without an accompanying shift in the downstream neurons receiving the changing feature label, such a change could disrupt the stability of the population code. However, our results suggest that population-level spatial representations remain stable even as the tuning of the underlying vRFs is shifting. In fact, spatial representations are even enhanced as a result of RF shifts.

An alternate proposal to a labeled line code relies on the joint information encoded across a population of cells (Erickson, 1982; Doetsch, 2000). This may occur at several scales–for example, V2 could use the pattern of information from V1 inputs to form a visual representation. This idea is more akin to an encoder-decoder model in which the downstream decoder does not need information about the altered representations in each of the encoder units, but instead relies on a population readout rule (Seriès et al., 2009). The population readout rule could incorporate knowledge about the 'labels' of the encoder units, but could perform equally well by relying on relative changes in the pattern across units to resolve the information encoded in the population. This may be a more parsimonious account of the attentional data reported so far. However, further exploration of population readout rules in visual cortex are needed to test this hypothesis.

Conclusions

The spatial encoding properties of the visual system can be measured and modeled at many different spatial scales. Here, we report these properties and how they change with attention for single voxels and for a group of voxels in a retinotopic region. Future lines of research into how attention modifies the specific inputs or outputs to a single encoding unit or a population of encoding units may help resolve the question of how shifts in RF labels are generated. Moreover, further investigation into population code readout rules may help adjudicate theories of sensory information processing beyond simple 'labeled line' coding schemes.

REFERENCES

Amano K, Wandell B a, Dumoulin SO (2009) Visual field maps, population receptive field sizes, and visual field coverage in the human MT+ complex. J Neurophysiol 102:2704–2718.

Anton-Erxleben K, Carrasco M (2013) Attentional enhancement of spatial resolution: linking behavioural and neurophysiological evidence. Nat Rev Neurosci 14:188–200.

Anton-Erxleben K, Stephan VM, Treue S (2009) Attention reshapes center-surround receptive field structure in macaque cortical area MT. Cereb Cortex 19:2466–2478.

Barlow HB (1972) Single units and sensation: A neuron doctrine for perceptual psychology? Perception 1:371–394.

Baruch O, Yeshurun Y (2014) Attentional attraction of receptive fields can explain spatial and temporal effects of attention. Vis cogn 22:704–736.

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. Ann Stat 29:1165–1188.

Brouwer GJ, Heeger DJ (2009) Decoding and reconstructing color from responses in human visual cortex. J Neurosci 29:13992–14003.

Brouwer GJ, Heeger DJ (2013) Categorical clustering of the neural representation of color. J Neurosci 33:15454–15465.

Butts DA, Goldman MS (2006) Tuning Curves, Neuronal Variability, and Sensory Coding. PLoS Biol 4:e92.

Cohen MR, Maunsell JHR (2014) Neuronal mechanisms of spatial attention in visual cerebral cortex. In: The Oxford Handbook of Attention, pp 318–345.

Connor CE, Gallant JL, Preddie DC, Van Essen DC (1996) Responses in Area V4 Depend on the Spatial Relationship Between Stimulus and Attention. J Neurophysiol 75:1306–1308.

Connor CE, Preddie DC, Gallant JL, Van Essen DC (1997) Spatial attention effects in macaque area V4. J Neurosci 17:3201–3214.

David S V, Hayden BY, Mazer JA, Gallant JL (2008) Attention to stimulus features shifts spectral tuning of V4 neurons during natural vision. Neuron 59:509–521.

Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Annu Rev Neurosci 18:193–222.

Dhruv NT, Carandini M (2014) Cascaded Effects of Spatial Adaptation in the Early Visual System. Neuron 81:529–535.

Doetsch GS (2000) Patterns in the brain. Neuronal population coding in the somatosensory system. Physiol Behav 69:187–201.

Dumoulin SO, Wandell BA (2008) Population receptive field estimates in human visual cortex. Neuroimage 39:647–660.

Erickson RP (1982) The across-fiber pattern theory: An organizing principle for molar neural function. In: Contributions to sensory physiology, Vol. 6, pp 79–110.

Ester EF, Anderson DE, Serences JT, Awh E (2013) A Neural Measure of Precision in Visual Working Memory. J Cogn Neurosci 25:754–761.

Ester EF, Sprague TC, Serences JT (2015) Parietal and Frontal Cortex Encode Stimulus-Specific Mnemonic Representations during Visual Working Memory. Neuron 87:893–905.

Gattass R, Nascimento-Silva S, Soares JGM, Lima B, Jansen AK, Diogo ACM, Farias MF, Botelho MMEP, Mariani OS, Azzi J, Fiorani M (2005) Cortical visual areas in monkeys: location, topography, connections, columns, plasticity and cortical dynamics. Philos Trans R Soc Lond B Biol Sci 360:709–731.

Harvey BM, Dumoulin SO (2011) The relationship between cortical magnification factor and population receptive field size in human visual cortex: constancies in cortical architecture. J Neurosci 31:13604–13612.

Huth AG, Nishimoto S, Vu AT, Gallant JL (2012) A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron 76:1210–1224.

Kay KN, Weiner KS, Grill-Spector K (2015) Attention reduces spatial uncertainty in human ventral temporal cortex. Curr Biol 25:595–600.

Klein BP, Harvey BM, Dumoulin SO (2014) Attraction of Position Preference by Spatial Attention throughout Human Visual Cortex. Neuron:1–11.

Lee S, Papanikolaou A, Logothetis NK, Smirnakis SM, Keliris G a (2013) A new method for estimating population receptive field topography in visual cortex. Neuroimage 81:144–157.

McAdams CJ, Maunsell JHR (1999) Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. J Neurosci 19:431–441.

Moran J, Desimone R (1985) Selective attention gates visual processing in the extrastriate cortex. Science (80-) 229:782–784.

Navalpakkam V, Itti L (2007) Search Goal Tunes Visual Features Optimally. Neuron 53:605–617.

Niebergall R, Khayat PS, Treue S, Martinez-Trujillo JC (2011) Expansion of MT neurons excitatory receptive fields during covert attentive tracking. J Neurosci 31:15499–15510.

Pouget A, Dayan P, Zemel RS (2003) Inference and computation with population codes. Annu Rev Neurosci 26:381–410.

Regan D, Beverley KI (1985) Postadaptation orientation discrimination. J Opt Soc Am 2:147–155.

Roberts M, Delicato LS, Herrero J, Gieselmann M a, Thiele A (2007) Attention alters spatial integration in macaque V1 in an eccentricity-dependent manner. Nat Neurosci 10:1483–1491.

Scolari M, Serences JT (2009) Adaptive Allocation of Attentional Gain. J Neurosci 29:11933–11942.

Scolari M, Serences JT (2010) Basing perceptual decisions on the most informative sensory neurons. J Neurophysiol 104:2266–2273.

Serences JT, Saproo S (2012) Computational advances towards linking BOLD and behavior. Neuropsychologia 50:435–446.

Seriès P, Stocker A a, Simoncelli EP (2009) Is the homunculus “aware” of sensory adaptation? Neural Comput 21:3271–3304.

Sheremata SL, Silver MA (2015) Hemisphere-Dependent Attentional Modulation of Human Parietal Visual Field Representations. J Neurosci 35:508–517.

Sprague TC, Ester EF, Serences JT (2014) Reconstructions of information in visual spatial working memory degrade with memory load. Curr Biol 24:2174–2180.

Sprague TC, Ester EF, Serences JT (2016) Restoring Latent Visual Working Memory Representations in Human Cortex. Neuron 91:694–707.

Sprague TC, Saproo S, Serences JT (2015) Visual attention mitigates information loss in small- and large-scale neural codes. Trends Cogn Sci:1–12.

Sprague TC, Serences JT (2013) Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. Nat Neurosci 16:1879–1887.

Tolias AS, Moore T, Smirnakis SM, Tehovnik EJ, Siapas AG, Schiller PH (2001) Eye movements modulate visual receptive fields of V4 neurons. Neuron 29:757–767.

Wandell BA, Winawer J (2015) Computational neuroimaging and population receptive fields. Trends Cogn Sci:1–9.

Womelsdorf T, Anton-Erxleben K, Pieper F, Treue S (2006) Dynamic shifts of visual receptive fields in cortical area MT by spatial attention. Nat Neurosci 9:1156–1160.

Womelsdorf T, Anton-Erxleben K, Treue S (2008) Receptive field shift and shrinkage in macaque middle temporal area through attentional gain modulation. J Neurosci 28:8934–8944.

Zirnsak M, Steinmetz N a, Noudoost B, Xu KZ, Moore T (2014) Visual space is compressed in prefrontal cortex before eye movements. Nature 507:504–507.

LEGENDS

Figure 1. Covert spatial attention task and hypothesized representation changes with shifts of spatial attention. (**a**) Subjects fixated centrally and attended to brief rotations in the pentagon stimulus on the left or right while a flickering checkerboard probe stimulus appeared at one of 51 grid locations across the visual field. On control runs, subjects attended to a contrast change at fixation. fMRI data measured during this attention task is used to create visualizable estimates of voxel receptive fields (vRFs) and stimulus reconstructions. (**b**) A receptive field model is fit to the responses of each voxel, and can be described by its x and y position (center), its response baseline, response amplitude, and its size (full-width half maximum). (**c**) Given a population of voxels in a retinotopic region, such as V1, we examine two different measures of spatial information in the population. The first, a spatial discriminability metric, scales with the slope of the tuning curve at a given location in space (**Materials and Methods**). The second relies on a multivariate inverted encoding model (IEM) for space. By reconstructing images of the mapping stimulus on each test trial, we can measure how population-level spatial information changes with attention. We then can model how changes in individual vRFs affect both of these population measures.

Figure 2. Changes in voxel receptive fields (vRFs) across attention conditions. We separately estimated vRFs for every voxel in visual and posterior parietal areas, discarding poorly estimated or noisy voxels (**Table 1**). Unless otherwise specified, figure data is averaged across subjects and error bars show 95% confidence intervals computed with resampling the data distribution. (**a**) An example vRF shows that attending covertly to the left location shifts the center of the receptive field profile to the left, when compared to the neutral attend fixation condition. Voxel is from subject AR in area V3A/B. (**b**) Our vRF estimates reproduced the canonical size-eccentricity relationship (positive slope in all ROIs, p < minimum possible p-value, e.g., 1/1000 iterations) and the increase in slope between visual regions. (**c**) Preferred position changes of V4 vRFs with covert spatial attention. We binned each vRF by its position during the attend fixation condition. The origin of each arrow is the center of each position bin. The end of the arrow shows the average position shift of the vRFs within that position bin during the attend peripheral conditions (left/right are collapsed and shown as attend left). The majority of vRFs shift toward the attended location (blue-green color map vs. red-yellow). (**d**) Mean changes in vRF parameters (attend peripheral target – attend fixation) in each visual area. (**e**) Attentional modulations of each vRF parameter plotted by the vRF's distance from the attention target. We only show areas where these data are significantly described by a polynomial of order n > 0 (**Table 2**).

Figure 3. Spatial discriminability increases with attention and is mediated by position changes in vRFs. Error bars depict bootstrapped 95% CIs. (**a**) We formulated a measurement to describe the ability of a population of voxels to make fine spatial discriminations around the attention target. We used the properties of each voxel's spatial tuning curve to make this measurement (**Materials and Methods**). Spatial discriminability increased when subjects attended the target, compared to when they ignored the target in the opposite hemifield (resampled p < minimum possible p-value (1/1000) for all ROIs for all ROIs). (**b**) The discriminability metric was recomputed for vRFs with a variety of attentional modulations. (*none* = vRF parameters during the neural attend fixation condition; *a* = amplitude; *s* = size; *p* = position). Spatial discriminability increased significantly when we applied position changes measured during the attend L/R task to the vRFs compared to when we applied no parameter changes (solid bar). By contrast, applying size changes either did not change spatial discriminability (V3A/B, V4, IPS0) or even decreased it from the no change condition (V1-V3).

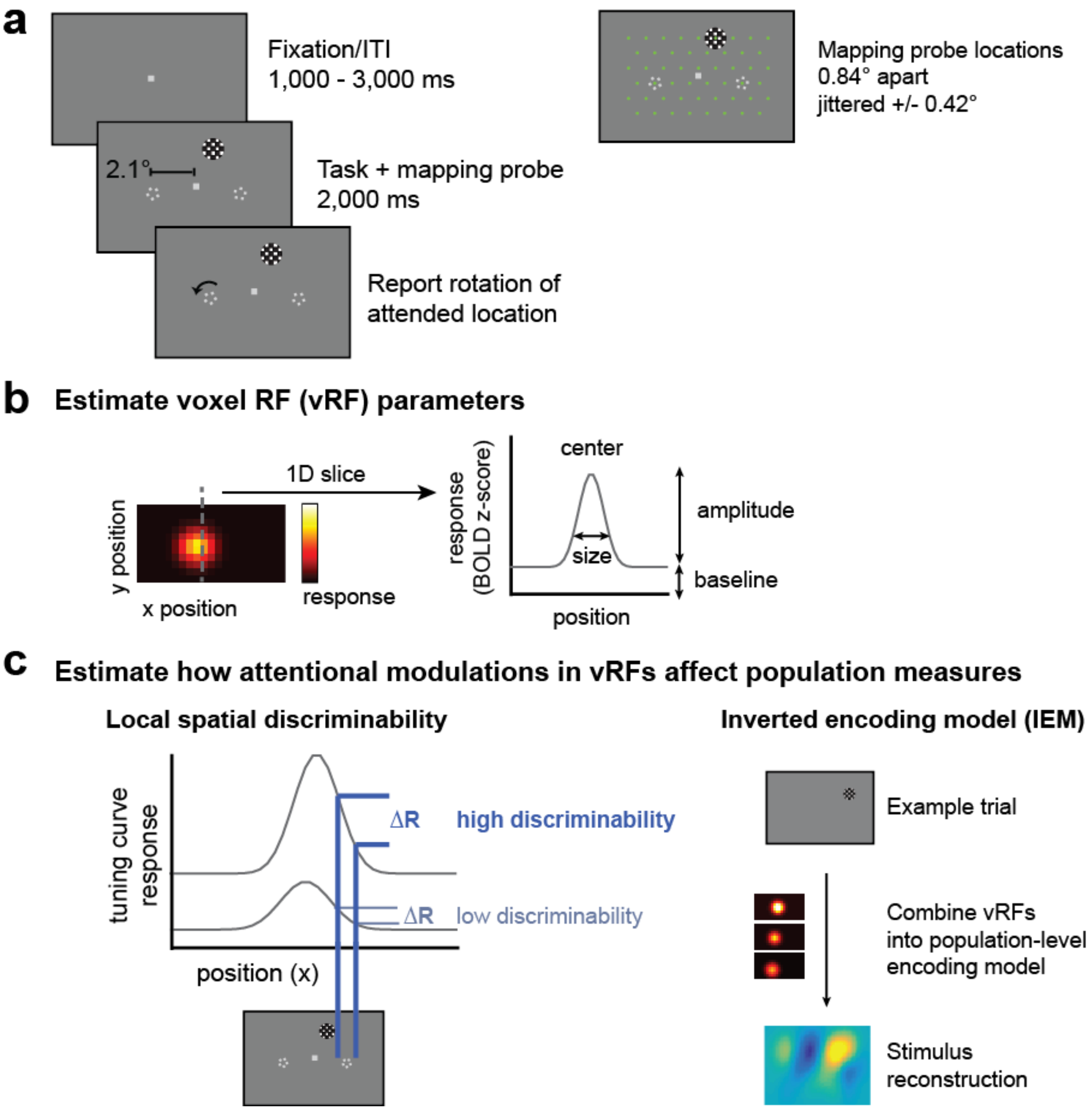
Figure 4. Multivariate inverted encoding model (IEM) used to reconstruct the mapping probe stimuli. (**a**) To train the IEM, we first take the BOLD data from all voxels within a visual region from a subset of training trials. Then, we solve for a set of channel weights using least squares regression. To reconstruct the stimulus, we invert this weight matrix and multiply it with BOLD data from the same voxels during a test trial. This yields a reconstructed channel response profile, which can be transformed into a reconstruction of the mapping stimulus on every trial in each attention condition. Data shown are examples from participant AR for a subset of V1 voxels. (**b**) Example stimulus reconstructions for participant AI, V1. These reconstructions were averaged across trials with the same position, yielding 51 reconstructions – one for each unique position in the test dataset. In the left panel, the same averaged position reconstructions are shown for each condition. The amplitude on the left is higher when attending left, and on the right when attending right. (**c**) Average reconstruction sizes and amplitudes for each stimulus position (collapsed across condition; left is attended). The diameter of the circle depicts the average fit FWHM of the reconstructions at that spatial position. Reconstruction amplitude was greater in the attended hemifield compared to the ignored hemifield in areas V3, V4, V3A/B, and IPS0, $p < 0.005$).

Figure 5. Reconstruction parameters as a function of mapping stimulus distance from the covertly attended locations (*s_dist_attn*) and attention hemifield (attended vs. ignored). See **Table 3** for complete list of p-values.

Figure 6. A layered spatial encoding model reveals how different sets of vRF changes lead to enhancements in multivariate stimulus reconstructions. (**a**) The first layer of the model uses the vRF fits to generate BOLD data from every subject's real trial sequence. Then the BOLD data from all voxels within one ROI is used to train a multivariate spatial encoding model and reconstruct the mapping stimuli. (**b**) Change in reconstruction amplitude in the attended vs. the ignored hemifield. Simulated reconstructions (black) qualitatively reproduce changes in reconstructions using real BOLD data from the same reduced set of voxels (gray bars). Furthermore, while position shifts in vRFs are necessary to observe increases in reconstruction amplitude near the attended location (blue), size changes are not (red). The results for the best model in each visual area is shown in yellow. (**c**) RMSE between each set of IEM fits and the full empirical dataset fits shown in **Fig 5**. The null baseline model (far left) is a layered IEM where the vRF parameters are the same across all attention conditions. We then added vRF attentional modulations for each parameter as shown in the matrix, with the full model on the far right. * indicate an FDR-corrected p-value <.05 for models that differed significantly from the null baseline model. The red bar highlights that a size change model generally performed significantly worse than the null. The green bar highlights that a position change model generally performed significantly better than the null.

921 **FIGURES**

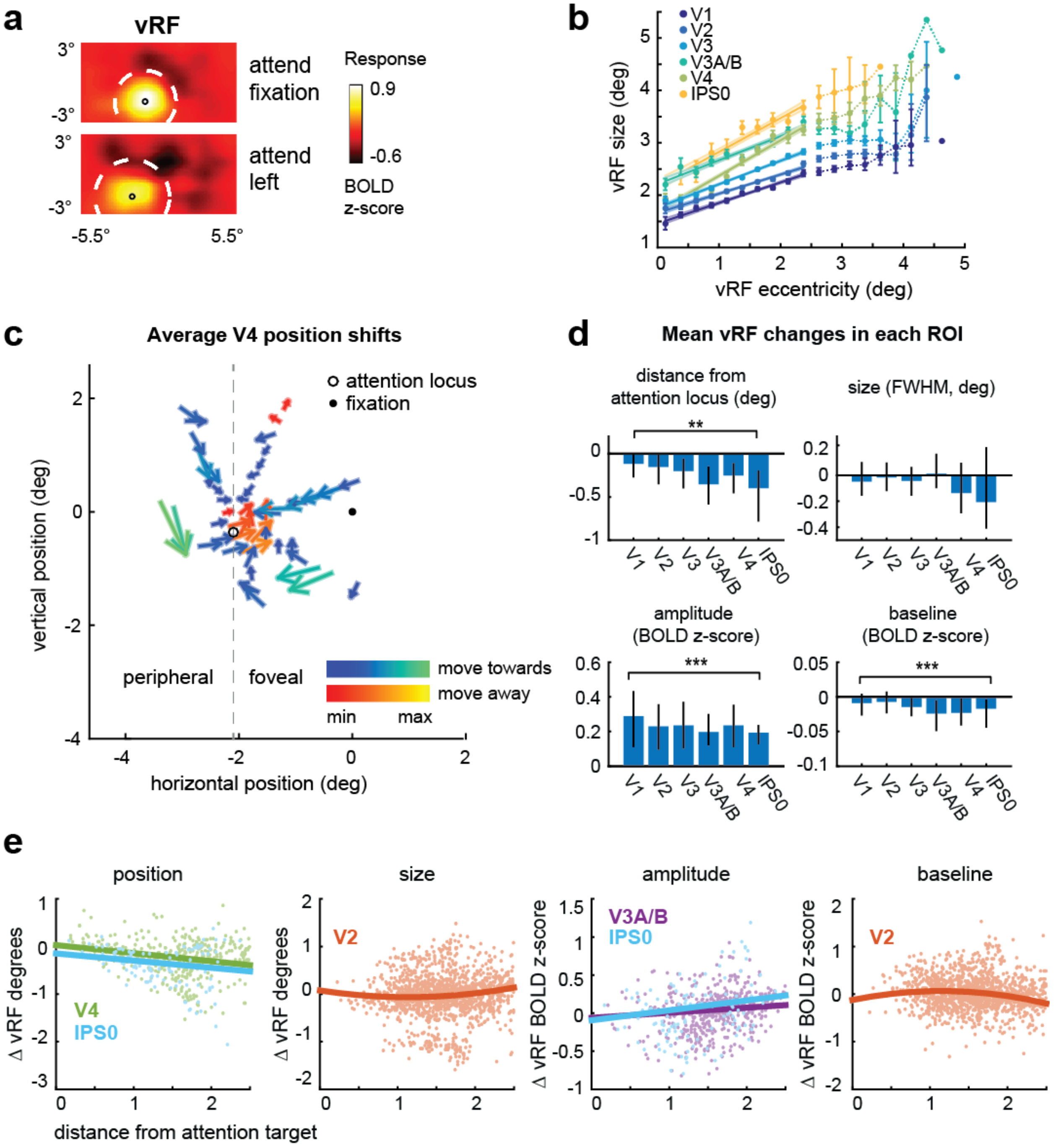
922 **Figure 1**



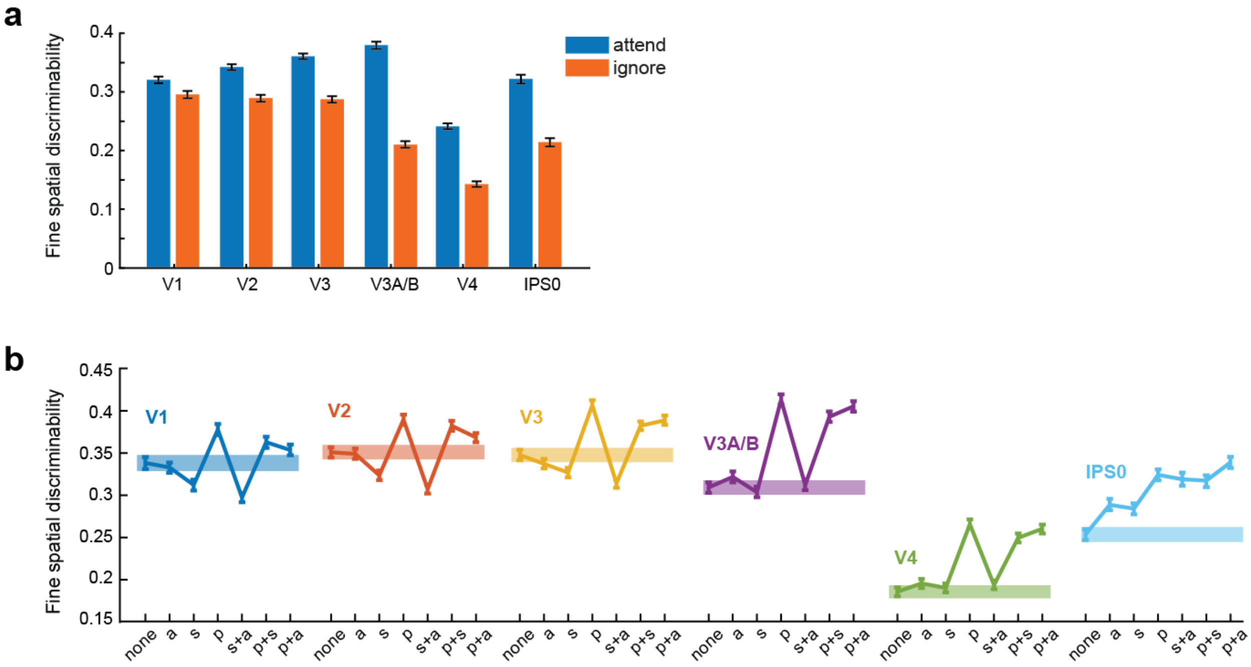
923

924

Figure 2

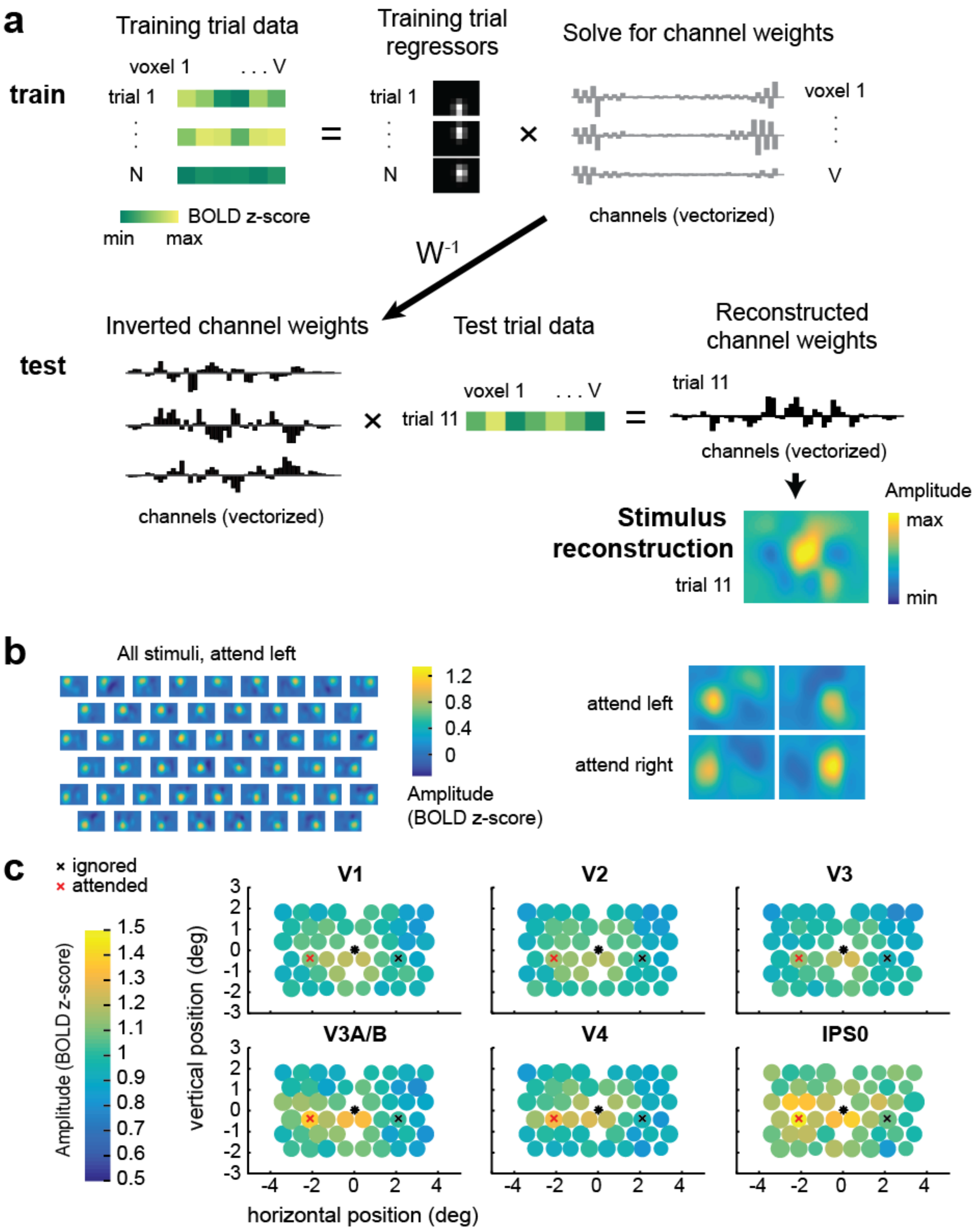

a
vRF
3°
-3°
3°
-3°
-5.5°
5.5°
attend fixation
attend left
Response
0.9
-0.6
BOLD z-score
b
V1
V2
V3
V3A/B
V4
IPS0
vRF size (deg)
vRF eccentricity (deg)
c
Average V4 position shifts
attention locus
fixation
vertical position (deg)
horizontal position (deg)
peripheral
foveal
move towards
move away
min
max
d
Mean vRF changes in each ROI
distance from attention locus (deg)
size (FWHM, deg)
amplitude (BOLD z-score)
baseline (BOLD z-score)
**
***
***
V1 V2 V3 V3A/B V4 IPS0
e
position
size
amplitude
baseline
Δ vRF degrees
Δ vRF BOLD z-score
V4
IPS0
V2
V3A/B
IPS0
V2
distance from attention target

928 **Figure 3**



929

930 **Figure 4**



931

932

Figure 5

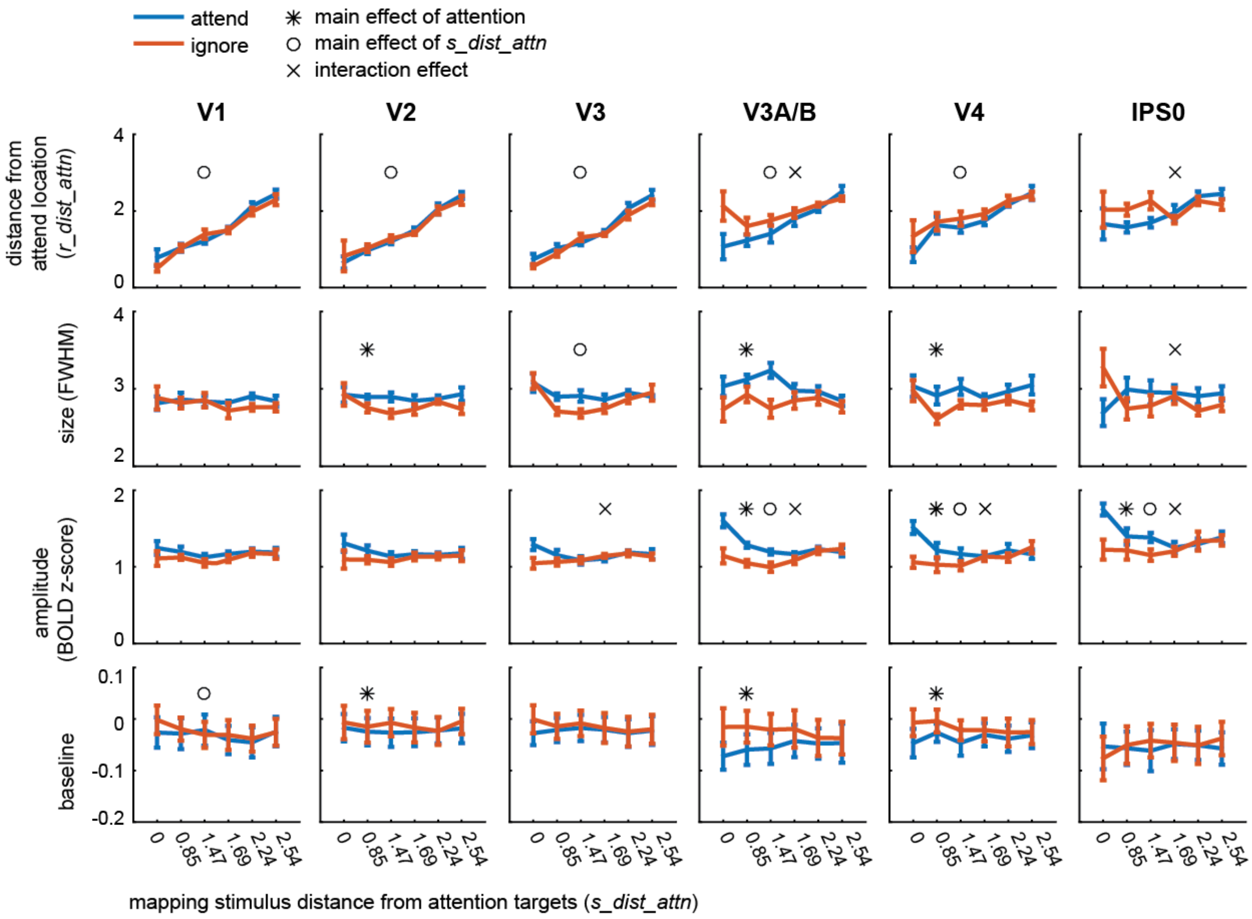
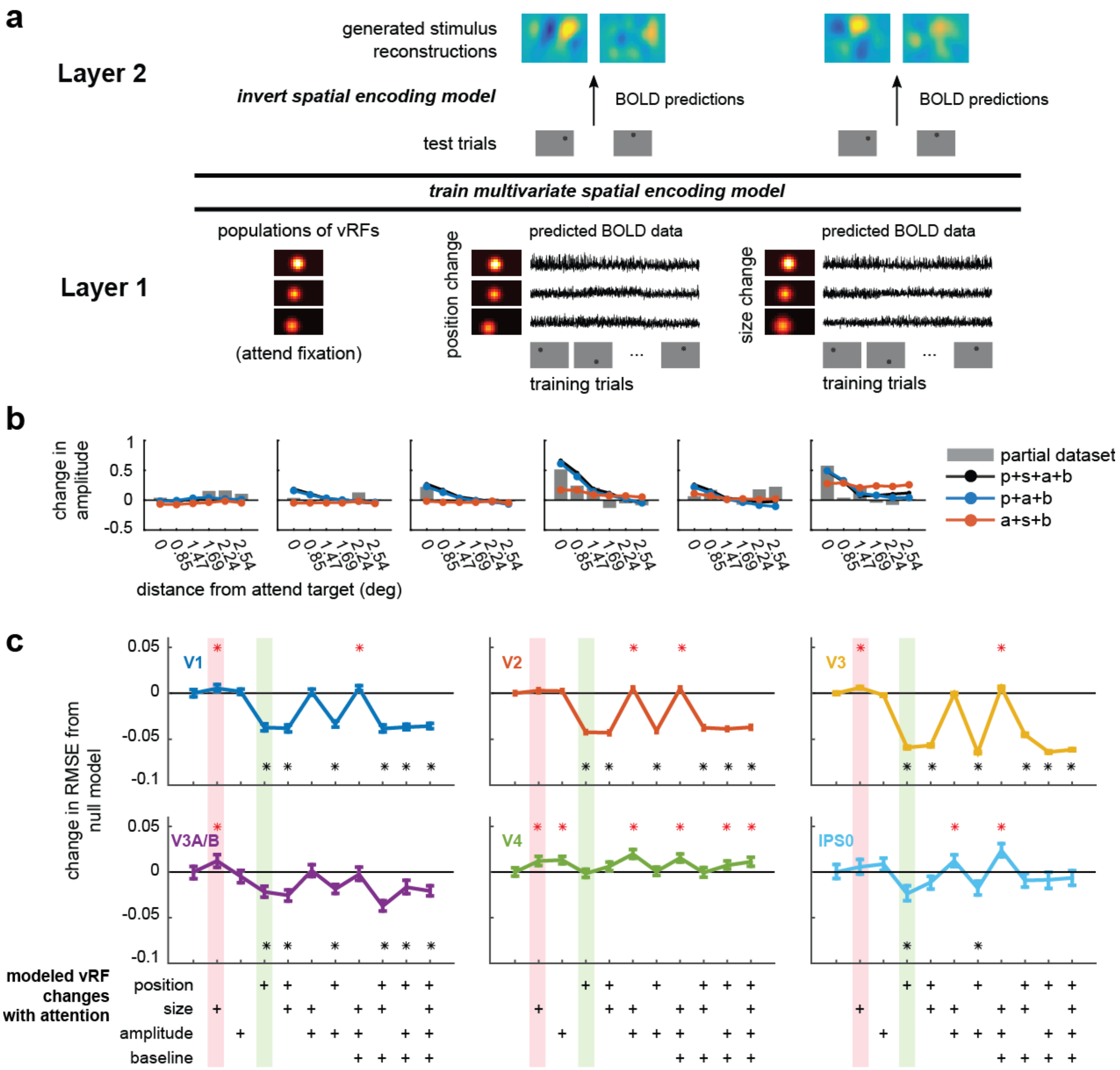

attend
ignore
✳ main effect of attention
○ main effect of *s_dist_attn*
× interaction effect
V1
V2
V3
V3A/B
V4
IPS0
distance from attend location (*r_dist_attn*)
size (FWHM)
amplitude (BOLD z-score)
baseline
0
2
4
3
1
0.1
-0.1
-0.2
0.85
1.47
1.69
2.24
2.54
mapping stimulus distance from attention targets (*s_dist_attn*)

Figure 6


a
Layer 2
generated stimulus reconstructions
invert spatial encoding model
BOLD predictions
BOLD predictions
test trials
train multivariate spatial encoding model
Layer 1
populations of vRFs
(attend fixation)
position change
predicted BOLD data
training trials
size change
predicted BOLD data
training trials
b
change in amplitude
1
0.5
0
-0.5
0 0.85 1.47 1.69 2.24 2.54
distance from attend target (deg)
partial dataset
p+s+a+b
p+a+b
a+s+b
c
change in RMSE from null model
0.05
0
-0.05
-0.1
V1
V2
V3
V3A/B
V4
IPS0
modeled vRF changes with attention
position
size
amplitude
baseline

TABLES

Table 1. vRF selection statistics, pooled across participants (N = 7)

Region of interest	Total number of localized voxels	Number of voxels after GLM thresholding	Number of voxels after regular-izability threshold	Number of voxels after cross-validation threshold	Percent that survive all thresholds	RMSE fit error for surviving voxels
V1	3,723	3,286	2,148	969	26.03%	0.1231
V2	4,154	3,685	2,895	1,355	32.62%	0.1228
V3	3,698	3,246	2,600	1,460	39.48%	0.1191
V3A/B	1,988	1,796	1,278	446	22.43%	0.1063
V4	1,702	1,308	954	349	20.51%	0.1060
IPS0	1,430	1,250	680	114	7.97%	0.0985
TOTAL	16,695	14,571	10,555	4,693	28.11%	0.1126

Table 2. Mean coefficients for polynomial fits of how vRF parameter change is modulated by distance from the attended location (*v_dist_attn*)

	Position	Size	Amplitude	Baseline
V1	.003	-.005, .032, -.023	-.015, -.003	.039, -.065
V2	-.022	*.109*, -.247, .075	-.031, .076, -.043	*-.143*, *.328*, -.123
V3	**-.103**	.007, -.015, -.025	.011, -.037	.128, -.221
V3A/B	**-.332**	*-.012*, -.017, .063	*.067*, -.061	-.122
V4	**-.168**, .028	-.056, .138, -.043	.002	-.039, .105, -.181
IPS0	*-.151*, **-.146**	-.080, .291, -.192	**.131, -.094**	*-.269*

[a] **bold** numbers indicate that the p-value passed FDR-correction (q = 0.05) across ROIs and coefficients within each parameter; *italicized* numbers are p < .05, uncorrected. Number of reported coefficients in the table correspond to the polynomial order which was yielded the most parsimonious fit to the data (e.g., 1 coefficient for n = 0, 2 coefficients for n = 1, etc.).

Table 3. 2-way ANOVA results for reconstruction parameter changes (*s_dist_attn* x attention hemifield).

	V1	V2	V3	V3A/B	V4	IPS0
Omnibus test						
Position	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**.007**
Size	.917	.097	**.001**	**.001**	**.017**	**.016**
Amplitude	.207	.220	**.003**	**<.001**	**<.001**	**<.001**
Baseline	.024	.257	.485	**.002**	**.004**	.925
Main effect of distance						
Position	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	.084
Size			**.001**	.233	.169	.679
Amplitude			.269	**<.001**	**.008**	**.007**
Baseline				.864	.336	
Main effect of attention						
Position	.573	.920	.399	.022	.189	.235
Size			.163	**.001**	**.005**	.509
Amplitude			.047	**.005**	**.002**	**.028**
Baseline				**<.001**	**.001**	
Interaction of distance & attention						
Position	.188	.892	.354	**.001**	.679	**.004**
Size			.157	.099	.582	**.005**
Amplitude			**.003**	**<.001**	**<.001**	**.002**
Baseline				.210	.202	

[a] **bold** numbers indicate that the p-value passed FDR-correction (q = .05, corrected across ROIs and comparisons within each parameter).

Table 4. RMSE (and 95% CIs) between reconstructions from the reduced dataset (only using voxels with RFs) or from different versions of the layered IEM using the same voxels.

	Real data	**p/s/a/b**	**p/a/b**	**p/s/b**	**s/a/b**	**p/a**	**s/a**	**p/s**	**p**	**a**	**s**	**none**
V1	0.773 [0.640, 0.913]	0.184 [0.181, 0.187]	0.183 [0.180, 0.186]	0.181 [0.177, 0.185]	0.225 [0.222, 0.228]	0.186 [0.183, 0.189]	0.221 [0.218, 0.224]	0.181 [0.177, 0.185]	0.182 [0.178, 0.186]	0.221 [0.218, 0.224]	0.225 [0.221, 0.229]	0.219 [0.216, 0.223]
V2	0.394 [0.326, 0.461]	0.185 [0.183, 0.187]	0.184 [0.182, 0.185]	0.185 [0.182, 0.187]	0.227 [0.226, 0.229]	0.181 [0.180, 0.183]	0.228 [0.226, 0.229]	0.179 [0.177, 0.182]	0.180 [0.178, 0.182]	0.225 [0.223, 0.226]	0.225 [0.223, 0.227]	0.222 [0.221, 0.224]
V3	0.368 [0.280, 0.480]	0.180 [0.179, 0.181]	0.177 [0.176, 0.179]	0.196 [0.194, 0.198]	0.247 [0.246, 0.249]	0.177 [0.176, 0.178]	0.241 [0.239, 0.242]	0.184 [0.183, 0.186]	0.182 [0.181, 0.184]	0.239 [0.238, 0.241]	0.247 [0.246, 0.249]	0.241 [0.240, 0.243]
V3A/B	0.745 [0.522, 0.978]	0.336 [0.331, 0.342]	0.340 [0.333, 0.348]	0.320 [0.314, 0.326]	0.354 [0.348, 0.362]	0.338 [0.334, 0.343]	0.359 [0.352, 0.365]	0.331 [0.325, 0.337]	0.335 [0.329, 0.342]	0.352 [0.345, 0.359]	0.369 [0.362, 0.376]	0.357 [0.350, 0.363]
V4	0.729 [0.617, 0.843]	0.393 [0.389, 0.398]	0.390 [0.386, 0.394]	0.381 [0.376, 0.387]	0.398 [0.393, 0.402]	0.383 [0.378, 0.388]	0.402 [0.397, 0.407]	0.388 [0.384, 0.393]	0.381 [0.376, 0.385]	0.395 [0.391, 0.399]	0.394 [0.389, 0.399]	0.382 [0.378, 0.387]
IPS0	0.879 [0.764, 1.006]	0.491 [0.483, 0.499]	0.489 [0.480, 0.498]	0.489 [0.481, 0.496]	0.521 [0.514, 0.529]	0.480 [0.473, 0.488]	0.510 [0.504, 0.517]	0.486 [0.479, 0.493]	0.474 [0.466, 0.483]	0.506 [0.500, 0.513]	0.503 [0.495, 0.512]	0.498 [0.491, 0.506]

[a] To generate CIs, the resampling of the real data is performed at the level of the fits to the reconstructions, whereas resampling layered IEM RMSEs is described in **Materials and Methods**