Dissociable signatures of visual salience and behavioral relevance across attentional priority maps in human cortex

Sprague TC, Itthipuripat S, Vo VA, Serences JT. Dissociable signatures of visual salience and behavioral relevance across attentional priority maps in human cortex. J Neurophysiol 119: 2153–2165, 2018. First published February 28, 2018; doi:10.1152/jn.00059.2018. 2018.—Computational models posit that visual attention is guided by activity within spatial maps that index the image-computable salience and the behavioral relevance of objects in the scene. These spatial maps are theorized to be instantiated as activation patterns across a series of retinotopic visual regions in occipital, parietal, and frontal cortex. Whereas previous research has identified sensitivity to either the behavioral relevance or the image-computable salience of different scene elements, the simultaneous influence of these factors on neural “attentional priority maps” in human cortex is not well understood. We tested the hypothesis that visual salience and behavioral relevance independently impact the activation profile across retinotopically organized cortical regions by quantifying attentional priority maps measured in human brains using functional MRI while participants attended one of two differentially salient stimuli. We found that the topography of activation in priority maps, as reflected in the modulation of region-level patterns of population activity, independently indexed the physical salience and behavioral relevance of each scene element. Moreover, salience strongly impacted activation patterns in early visual areas, whereas later visual areas were dominated by relevance. This suggests that prioritizing spatial locations relies on distributed neural codes containing graded representations of salience and relevance across the visual hierarchy.

INTRODUCTION

In a typical visual environment, some portions of the scene are visually salient by virtue of their image-computable properties, such as luminance, contrast, color saturation, and local feature contrast (Masciocchi et al. 2009; Parkhurst et al. 2002; Usher and Niebur 1996). Many computational models emphasize the importance of salience in guiding spatial attention (Itti and Koch 2001; Koch and Ullman 1985; Theeuwes 1994, 2010; Van der Stigchel et al. 2009; Veale et al. 2017; Wolfe 1994). However, visually salient portions of the scene do not necessarily contain behaviorally relevant information. For example, the flashing lights of a police car are relevant if they appear in the rearview mirror, but less so if they appear in oncoming traffic. Accordingly, other models emphasize behavioral relevance in guiding attention, with relevant objects overriding even highly salient scene elements (Egeth and Yantis 1997; Folk et al. 1992, 2002; Veale et al. 2017; Yantis and Johnson 1990; Zelinsky and Bisley 2015).

One unifying framework posits that both visual salience and behavioral relevance dissociably influence the activity profile across retinotopic cortex to compute an attentional priority map within each region, which can be read out to guide decisions and actions (Fig. 1; Fecteau and Munoz 2006; Itti and Koch 2001; Serences and Yantis 2006). This predicts that the topography of activation across an individual retinotopic map (e.g., V1) will reflect the spatial profile of visual salience and/or behavioral relevance of items throughout the scene, and each region’s priority map will weight these factors to varying degrees. Moreover, because representations of visual salience are more closely aligned with retinal input, we predict that attentional priority maps should transition from salience-driven to relevance-driven across the processing hierarchy.


© 2018 the American Physiological Society
Fig. 1. Identifying maps of visual salience and behavioral relevance. A: we define a salience map as a map of the visual scene where each position on the map indexes the importance of the corresponding location within the scene based on image-computable features, such as contrast, motion, or distinctness from the background or other scene items. Accordingly, activity within a salience map should scale with the visual salience of local scene elements. For an example scene, in which two stimuli of differing contrast are presented and a participant is cued to attend to one (dashed yellow circle), a salience map would show higher activation at visual field positions corresponding to higher contrast, even if those elements of the scene are not relevant for behavior. B: for a given location in the scene, activation in a pure salience map would scale only with image-computable features, such as contrast (shown). C: we define a relevance map as a map of the visual scene where each position on the map indexes the behavioral relevance of the corresponding location. In this example scene, the relevance map would only show activation at locations relevant to the behavior of the observer, independent of their visual salience. This requires that visual locations corresponding to highly salient but irrelevant stimuli are not reflected in a pure relevance map. D: a location within a pure relevance map would show high activity when the corresponding position is relevant for behavior and low activity when it is irrelevant. Importantly, a cortical map of visual space can reflect a combination of both visual salience and behavioral relevance by virtue of its position within the visual processing hierarchy.

Klein et al. 2014; Murray 2008; Pestilli et al. 2011; Serences and Yantis 2007; Sheremata and Silver 2015; Silver et al. 2005; Somers et al. 1999; Sprague and Serences 2013; Vo et al. 2017), and representations of behavioral relevance in the absence of sensory inputs during attention (Kastner et al. 1999) and working memory (Jerde et al. 2012; Rahmati et al. 2018; Saber et al. 2015; Sprague et al. 2014, 2016). Although these studies have demonstrated strong preliminary evidence for the existence of representations of attentional priority in retinotopic cortical regions, less is known about how these factors interact when both salience and relevance are parametrically manipulated. For instance, how is an extremely salient, yet behaviorally irrelevant distractor stimulus represented within region-level attentional priority maps spanning occipital, parietal, and frontal cortex? Additionally, many previous studies have required subjects to search for a target stimulus (e.g., letter or line) among an array of distractors, with one stimulus marked as salient either by virtue of a unique feature or by the abrupt onset of an array element (Balan and Gottlieb 2006; Bertie et al. 2016; Gottlieb et al. 1998; Ipatia et al. 2006; Poloratski et al. 2017; Thompson et al. 1997; Yildirim et al. 2018). In such visual search tasks, the participant explores the scene freely before making a response, and so the most relevant location is updated moment by moment, rather than being at a controlled location. Finally, previous single-unit neurophysiology experiments assessing the effects of visual salience and behavioral relevance on neural firing rates have highlighted the necessity of assaying the activity levels of all neurons within a region to infer properties of neural priority maps, as opposed to just measuring those neurons with receptive fields overlapping the relevant stimulus location (Bisley and Goldberg 2003).

In this study, we addressed these issues by parametrically manipulating the salience (luminance contrast) and relevance (attended or unattended) of two visual stimuli and by reconstructing neural “priority maps” from blood oxygen level-dependent (BOLD) activation profiles across entire retinotopic regions. By quantifying aspects of these maps, we could identify how salience and relevance interact to determine population-level representations of attentional priority (Fig. 1). We found that early visual areas (e.g., V1) were sensitive to salience, whereas both earlier and later visual areas (e.g., IPS0) were sensitive to relevance. These results demonstrate a transition from salience-dominated attentional priority maps in early visual cortex to representations dominated by behavioral relevance in higher stages of the visual system.

MATERIALS AND METHODS

Experimental design. We recruited eight participants (1 man, all right-handed, age 26.5 ± 1.15 yr, mean ± SE), including one author (subject AP). Two of these participants had never participated in visual functional neuroimaging experiments before (subjects BA and BF). All others had participated in other experiments in the laboratory (Ester et al. 2015; Sprague and Serences 2013; Sprague et al. 2014, 2016; Vo et al. 2017). All procedures were approved by the University of California, San Diego (UCSD) Institutional Review Board, all participants gave written informed consent before participating, and all participants were compensated for their time ($20/h for scanning sessions, $10/h for behavioral sessions; subject AP was not compensated).

Each participant performed a 1-h training session before scanning during which they were familiarized with all tasks performed inside the scanner. We also used this session to establish initial behavioral performance thresholds by manipulating task difficulty across behavioral blocks.

We scanned participants for a single 2-h main task scanning session comprising at least 4 mapping task runs, used to independently estimate encoding models for each voxel (see Fig. 3A), and 4 selective attention task runs (broken into 2 sub-runs each, see below; used to independently select voxels for inclusion in all analyses). All participants also underwent additional localizer and retinotopic mapping scanning sessions to independently identify regions of interest (ROIs; see Region of interest definition).

We presented stimuli using the Psychophysics toolbox (Brainard 1997; Pelli 1997) for MATLAB (The MathWorks, Natick, MA). During scanning sessions, we rear-projected visual stimuli onto a 110-cm-wide screen placed ~370 cm from the participant’s eyes at the foot of the scanner bore using a contrast-linearized liquid crystal display (LCD) projector (1,024 × 768, 60 Hz). In the behavioral familiarization session, we presented stimuli on a contrast-linearized LCD monitor (1,920 × 1,080, 60 Hz) 62 cm from participants, who were comfortably seated in a dimmed room and positioned using a chin rest. For all sessions and tasks (main selective attention task, mapping task, and localizer), we presented all stimuli on a neutral gray 6.82° radius circular aperture, surrounded by black (only aperture shown in Fig. 2A).

Selective attention task. We instructed participants to attend to one of two random line stimuli (RLS), each appearing at one of three contrasts (20%, 40%, 80%). On each trial, each RLS stimulus would cohere into a “spiral” form, and participants responded with one of two button presses indicating which direction of spiral they detected
stimulus. The entire stimulus display was rotated randomly about fixation on each trial, so each and every trial involved a distinct stimulus display. Each trial, 2 stimuli appeared at 3.5° eccentricity and could be either 72° or 144° polar angle apart (CW from radial) (see left inset). Targets (1,000 ms) were variable-coherence spirals called “Type 1” (line angle CCW from radial) or Type 2 (shown: line angle CW from radial) (see right inset). Participants responded with a button press to indicate which type of spiral appeared in the cued stimulus. ITI intertrial interval. B: we presented the attended stimulus (yellow dashed circle) and unattended stimuli each at 1 of 3 logarithmically spaced contrast levels, fully crossed. C: on each trial, 2 stimuli appeared at 3.5° eccentricity and could be either 72° or 144° polar angle apart (inset). The attended stimulus (yellow) for this example arrangement could be any of the 5 “base” positions. The attended stimulus could appear CW (solid lines) or CCW (dashed lines) with respect to the unattended stimulus. The entire stimulus display was rotated randomly about fixation on each trial, so each and every trial involved a distinct stimulus display. D and E: behavioral performance (D) and response time (E) did not reliably vary as a function of attended or unattended stimulus contrast, or their interaction (2-way permuted repeated-measures ANOVA, P values for main effect of attended contrast, unattended contrast, and interaction for accuracy: 0.359, 0.096, 0.853; for response time: 0.926, 0.143, 0.705).

Both stimuli appeared along an invisible iso-eccentric ring in the cued RLS. We designed this stimulus to minimize the allocation of non-spatial feature-based attention, as well as to minimize the influence of potential radial biases in orientation preference in visual cortex as a function of preferred polar angle (Freeman et al. 2011, 2013).

Each trial began with a 500-ms attention cue at fixation indicating which of the two stimuli to attend, followed by appearance of both stimuli simultaneously. Because this task required participants to identify the locations of both stimuli and then select one on the basis of the color of the attention cue, we chose contrasts on the basis of pilot testing such that participants could easily localize even the lowest contrast stimulus. Moreover, other types of stimuli presented at these contrasts (Gabor s and/or gratings) result in dynamic responses across early visual cortex in human functional MRI (fMRI; Buracas and Boynton 2007; Gandhi et al. 1999; Gouws et al. 2014; Murray 2008; Pestilli et al. 2011). Although it is possible that responses in other regions, such as the retinotopic intraparietal sulcus (IPS) regions, could saturate at much lower contrasts, using lower stimulus contrast and the purely endogenous, nonspatial attention cue would have resulted in participants attending to the incorrect location on some trials when one of the two stimuli was low contrast. Both stimuli remained onscreen for 3,000 ms, during which time a 1,000-ms spiral target appeared independently and instantaneously at each stimulus position. The target onset was randomly chosen on each trial for each stimulus (attended and unattended) from a uniform distribution spanning 500–1,500 ms.

Both stimuli always appeared along an invisible iso-eccentric ring 3.5° from fixation and had a radius of 1.05°. On each trial, the two stimuli appeared either 72° or 144° polar angle apart (4.11° and 6.66° distance between centers, respectively). We randomly rotated the stimulus array on each trial (0–72° polar angle) around fixation, so the positions of the stimuli on each trial were entirely unique. The color of the attention cue presented at fixation indicated with 100% validity whether to attend to the clockwise (blue) or the counterclockwise (red) stimulus (Fig. 2A). Each trial was separated by an intertrial interval (ITI) drawn from a uniform distribution spanning 2.5–6.5 s at 0.09-s steps [distribution of ITIs was computed with MATLAB’s linspace command as ITIs = linspace(2.5,6.5,45)], resulting in an average trial duration of 8 s.

Both stimuli flickered in-phase at 15 Hz (2 frames on, 2 frames off, monitor refresh rate of 60 Hz). Each stimulus consisted of 35 light and 35 dark lines, each 0.3° long and 0.035° thick, that were replotted during each flicker period centered at random coordinates drawn from a uniform disk 0.9° in radius. On flicker periods with no targets, the orientation of each line was drawn from a uniform distribution. On flicker periods with targets, the orientation of a random subset of lines (defined by the target coherence) was either incremented (i.e., adjusted counterclockwise, “Type 1 spiral”) or decremented (i.e., adjusted clockwise, “Type 2 spiral”) from the Cartesian angle of each line relative to stimulus center (see Fig. 2A). Participants responded with a left button press to report a Type 1 spiral target or a right button press to indicate a Type 2 spiral target.

We counterbalanced the attended and unattended stimuli on the basis of contrast (20%, 40%, or 80%; Fig. 2B), stimulus separation distance (72° or 144° polar angle; Fig. 2C), and approximate position of the attended stimulus (1 of 5 “base” positions). Accordingly, for a single repetition of all trial types, we acquired 90 trials, broken up into 2 sub-runs, each lasting 382 s (45 trials, 8 s each, 12-s blank screen at beginning of scan, 10-s blank screen at end of each scan). Participants
performed 8 sub-runs, resulting in a full data set of 360 trials per participant.

To keep performance below ceiling and at an approximately fixed level (at ~80%), we adjusted the coherence of targets (defined as the percentage of lines forming a spiral target on each flicker cycle; Fig. 2F) independently for 20%, 40%, and 80% contrast targets before the start of each full run (i.e., 2 sub-runs).

Spatial mapping task. To estimate a spatial encoding model for each voxel (see below), we presented a flickering checkerboard stimulus at different positions across the screen on each trial. Participants attended these stimuli to identify rare target events [changes in checkerboard contrast on 10 of 47 (21.2%) of trials, evenly split between increments and decrements]. During each run, we chose the position of each trial’s checkerboard (0.9° radius, 70% contrast, 6-Hz full-field flicker) from a triangular grid of 37 possible positions and added a random uniform circular jitter (0.5° radius; Fig. 2C). As in a previous report, we rotated the base position of the triangular grid on each scanner run to increase the spatial sampling density (Sprague et al. 2016). Accordingly, every mapping trial was unique. The base triangular grid of stimulus positions separated stimuli by 1.5° and extended 4.5° from fixation (3 steps). This, combined with random jitter and the radius of the mapping stimulus, resulted in a visual “field of view” (region of the visual field stimulated) of 5.9° from fixation for our spatial encoding model. On trials in which the checkerboard stimulus overlapped the fixation point, we drew a small aperture around fixation (0.8° diameter).

Each trial consisted of a 3,000-ms stimulus presentation period followed by a 2,000- to 6,000-ms ITI (uniformly sampled). On trials with targets, the checkerboard was dimmed or brightened for 500 ms, beginning at least 500 ms after stimulus onset and ending at least 500 ms before stimulus offset. We instructed participants to only respond if they detected a change in checkerboard contrast and to minimize false alarms. We discarded all target-present trials when estimating spatial encoding models. To ensure participants performed below ceiling, we adjusted the task difficulty between each mapping run by changing the percent contrast change for target trials. Each run consisted of 47 trials (10 of which included targets), a 12-s blank period at the beginning of the run, and a 10-s blank period at the end of the run, totaling 352 s.

Visual attention localizer task. To identify voxels responsive to the region of the screen subtended by the mapping and selective attention stimuli, all participants performed several runs of a visual localizer task reported previously (Sprague et al. 2016). Participants performed between three and eight runs of this task in total. For one participant, we used data from the same task, acquired for a different experiment (Sprague et al. 2016) at a different scanning resolution (subject AS, 2 × 2 × 3-mm³ voxel size), resampled to the resolution used in the present study. For four participants, including the one scanned with ASSET, a 12-s blank period at the beginning of the run, and a 10-s blank period at the end of the run, totaling 352 s.

Each trial consisted of a flickering radial checkerboard hemi-annulus presented on the left or right half of fixation, subtending 0.8° to 6.0° eccentricity around fixation (6-Hz contrast reversal flicker, 100% contrast, 10.0-s duration). On each trial, participants performed a demanding spatial working memory (WM) task in which they carefully maintained the position of a target stimulus (red dot presented over the stimulus, 500 ms) over a 3,000-ms delay interval, after which a probe stimulus (green dot, 750 ms) appeared near the remembered position. Participants reported whether the green probe dot appeared to the left or right of the remembered position, or above or below the remembered position, as prompted by the appearance of a 1.0°-long bar at fixation (horizontal bar: left vs. right; vertical bar: above vs. below; 1.5-s response window). We manipulated the target-probe separation distance between runs to ensure performance was below ceiling.

WM trials could occur beginning 1.0 s after checkerboard onset and ended at latest 2.5 s before checkerboard offset. All trials were separated by a 3- to 5-s ITI (uniformly spaced across trials), and we included 4 null trials in which no checkerboard or WM stimuli appeared (10 s long each). Each run featured 16 total stimulus-present trials, a 14-s blank screen at the beginning of the run, and a 10-s blank screen at the end of the run, totaling 304 s.

fMRI acquisition. We scanned all participants on a 3T research-dedicated GE MR750 scanner located at the UCSD Keck Center for Functional Magnetic Resonance Imaging with a 32-channel send/receive head coil (Nova Medical, Wilmington, MA). We acquired functional data using a gradient echo-planar imaging (EPI) pulse sequence (19.2 × 19.2-cm² field of view, 64 × 64 matrix size, 35 3-mm-thick slices with 0-mm gap, axial orientation, TR = 2,000 ms, TE = 30 ms, flip angle = 90°, voxel size = 3 mm isotropic).

To anatomically coregister images across sessions and within each session, we also acquired a high-resolution anatomical scan during each scanning session (fast spoiled gradient-echo T1-weighted sequence, TR/TE = 11/3.5 ms, TI = 1,100 ms, 172 slices, flip angle = 18°, resolution = 1 mm³). For all sessions but one, anatomical scans were acquired with ASSET (array coil spatial sensitivity encoding) acceleration. For the remaining session, we used an 8-channel send/receive head coil and no ASSET acceleration to acquire anatomical images with minimal signal inhomogeneity near the coil surface, which enabled improved segmentation of the gray-white matter boundary. We transformed these anatomical images to Talairach space and then reconstructed the gray-white matter surface boundary in BrainVoyager 2.6.1 (Brain Innovation, Maastricht, The Netherlands), which we used for identifying ROIs.

fMRI preprocessing. We preprocessed fMRI data as described in our previous reports (Sprague et al. 2014, 2016). We coregistered functional images to a common anatomical scan across sessions (used to identify gray-white matter surface boundary as described above) by first aligning all functional images within a session to that session’s anatomical scan, and then aligning that session’s scan to the common anatomical scan. We performed all preprocessing using FSL (Oxford, UK) and BrainVoyager 2.6.1 (Brain Innovation). Preprocessing included unwarping the EPI images using routines provided by FSL, then slice-time correction, three-dimensional motion correction (6-parameter affine transform), temporal high-pass filtering (to remove first-, second-, and third-order drift), transformation to Talairach space (resampling to 3 × 3 × 3-mm³ resolution) in BrainVoyager, and, finally, normalization of signal amplitudes by converting to Z scores separately for each run using custom MATLAB scripts. We did not perform any spatial smoothing beyond the smoothing introduced by resampling during the coregistration of the functional images, motion correction, and transformation to Talairach space. All subsequent analyses were computed using custom code written in MATLAB (release 2015a).

Region of interest definition. On the basis of our previous work (Sprague et al. 2014, 2016), we identified 10 a priori ROIs using independent scanning runs from those used for all analyses reported in the text. For retinotopic ROIs (V1–V3, hV4, V3a, IPS0–IPS3), we utilized a combination of retinotopic mapping techniques. Each participant completed several scans of meridian mapping in which we alternately presented flickering checkerboard “bowties” along the horizontal and vertical meridians. Additionally, each participant completed several runs of an attention-demanding polar angle mapping task in which they detected brief contrast changes of a slowly rotating checkerboard wedge (described in detail in Sprague and Serences 2013). We used a combination of maps of visual field meridians and polar angle preference for each voxel to identify retinotopic ROIs (Engel et al. 1994; Swisher et al. 2007). Polar angle maps computed using the attention-demanding mapping task for several participants are available in previous publications (participant AI, Sprague and Serences 2013; participants AL and AP, Ester et al. 2015). We combined left and right hemispheres for all ROIs, as well as dorsal and ventral aspects of V2 and V3 for all analyses, by concatenating voxels.
We defined superior precentral sulcus (sPCS), the putative human homolog of macaque frontal eye field (FEF; Mackey et al. 2017; Srimal and Curtis 2008) by plotting voxels active during either the left or right conditions of the localizer task described above [false discovery rate (FDR) corrected, \( q = 0.05 \)] on the reconstructed gray-white matter boundary of each participant’s brain and manually identifying clusters appearing near the superior portion of the precentral sulcus, following previous reports (Srimal and Curtis 2008). Though retinotopic maps have previously been reported in this region (Hagler and Sereno 2006; Mackey et al. 2017), we did not observe topographic visual field representations. We anticipate this due to the substantially limited visual field of view we could achieve inside the scanner (maximum eccentricity: \( \sim 7^\circ \)).

Inverted encoding model. To reconstruct images of salience and/or relevance maps carried by activation patterns measured over entire ROIs, we implemented an inverted encoding model (IEM) for spatial position (Sprague and Serences 2013). This analysis involves first estimating an encoding model [sensitivity profile over the relevant feature dimension \( s \) parameterized by a small number of modeled information channels] for each voxel in a region by using a "training set" of data reserved for this purpose (4 spatial mapping runs). The encoding models across all voxels within a region are then inverted to estimate a mapping used to transform novel activation patterns from a "test set" (selective attention task runs) into activation in the modeled set of information channels.

Adopting analysis procedures from previous work, we built an encoding model for spatial position based on a linear combination of spatial filters (Sprague and Serences 2013; Sprague et al. 2014, 2015). Each voxel’s response was modeled as a weighted sum of 37 identically shaped spatial filters arrayed in a triangular grid (Fig. 3). Centers were spaced by 1.59°, and each filter was a Gaussian-like function with full-width half-maximum of 1.75°:

\[
f(r) = \left( 0.5 + 0.5 \cos \left( \frac{\pi r}{s} \right) \right)^7 \quad (1)
\]

for \( r < s \); 0 otherwise,

where \( r \) is the distance from the filter center and \( s \) is a "size constant" reflecting the distance from the center of each spatial filter at which the filter returns to 0. Values greater than this are set to 0, resulting in a single smooth round filter at each position along the triangular grid (\( s = 4.404^\circ \); see Fig. 3 for illustration of filter layout and shape; see also Sprague and Serences 2013; Sprague et al. 2014, 2016).

This triangular grid of filters forms the set of information channels for our analysis. Each mapping task stimulus is converted from a contrast mask (1 for each pixel subtended by the stimulus, 0 elsewhere) to a set of filter activation levels by taking the dot product of the vectorized stimulus mask and the sensitivity profile of each filter. This results in each mapping stimulus being described by 37 filter activation levels rather than \( 1,024 \times 768 = 786,432 \) pixel values. Once all filter activation levels are estimated, we normalize so that the maximum filter activation is 1.

Following previous reports (Brouwer and Heeger 2009; Sprague and Serences 2013), we model the response in each voxel as a weighted sum of filter responses (which can loosely be considered as hypothetical discrete neural populations, each with spatial receptive fields (RFs) centered at the corresponding filter position):

\[
B_k = C_k W
\]

where \( B_k \) (\( n \) trials \( \times \) \( m \) voxels) is the observed BOLD activation level of each voxel during the spatial mapping task (averaged over two TRs, 6.00—8.00 s after mapping stimulus onset), \( C_k \) (\( n \) trials \( \times \) \( k \) channels) is the modeled response of each spatial filter, or information channel, on each non-target trial of the mapping task (normalized from 0 to 1), and \( W \) is a weight matrix (\( k \) channels \( \times \) \( m \) voxels) quantifying the contribution of each information channel to each voxel. Because we have more stimulus positions than modeled information channels, we can solve for \( W \) using ordinary least-squares linear regression:

\[
\hat{W} = (C_k^T C_k)^{-1} C_k^T B_k
\]

This step is univariate and can be computed for each voxel in a region independently. Next, we used all estimated voxel encoding

---

Fig. 3. Measuring salience and relevance maps with an inverted encoding model. A: we estimated a spatial encoding model using data from a spatial mapping task in which participants viewed flickering checkerboard disks presented across a triangular grid of positions spanning a hexagonal region of the screen with spatial jitter (see Sprague et al. 2016). We modeled each voxel’s spatial sensitivity (i.e., receptive field, RF) as a weighted sum of smooth functions centered at each point in a triangular grid spanning the stimulus display visible inside the scanner. By knowing the stimulus on each trial (contrast mask) and the sensitivity image alignment and coregistration.

B: we then used all voxels within a region to compute an inverted encoding model (IEM). This step is multivariate; all voxels contribute to the IEM. This IEM allows us to transform activation patterns measured during the main spatial attention task (Fig. 2A) to activations of modeled information channels. We then compute a weighted sum of information channels on the basis of their computed activation in the spatial attention task. These images of the visual field index the activation across the entire cortical region transformed into visual field coordinates. To quantify activation in the map, we extract the mean signal over the map area corresponding to the known stimulus positions on that trial (red dashed circle: attended stimulus position; black dashed circle: unattended stimulus position). For visualization of spatial maps averaged across trials, we rotated reconstructions as though stimuli were presented at positions indicated in cartoons (e.g., Fig. 2C). See Sprague et al. (2016) for detailed methods on image alignment and coregistration.
models within a ROI ($\hat{W}$) and a novel pattern of activation from the WM task (each TR from each trial, in turn) to compute an estimate of the activation of each channel ($\hat{C}_2$; $n$ trials $\times k$ channels), which gave rise to that observed activation pattern across all voxels within that ROI ($B_2$; $n$ trials $\times m$ voxels):

$$\hat{C}_2 = B_2 \hat{W}^T \left( W \hat{W}^T \right)^{-1} \tag{4}$$

Once channel activation patterns are computed (Eq. 4), we compute spatial reconstructions by weighting each filter’s spatial profile by the corresponding channel’s reconstructed activation level and summing all weighted filters together. This step aids in visualization, quantification, and coregistration of trials across stimulus positions but does not confer additional information.

We used four mapping task runs to estimate the encoding model for each voxel. We then inverted that encoding model to reconstruct visual field images during all main spatial attention task runs.

Because stimulus positions were unique on each trial of the selective attention task (Fig. 2C), direct comparison of image reconstructions on each trial is not possible without coregistration of reconstructions so that stimuli appeared at common positions across trials. To accomplish this, we adjusted the center position of the spatial filters on each trial such that we could rotate the resulting reconstruction. For Fig. 4, we rotated each trial such that one target (the unattended stimulus, Fig. 2B) was centered at $x = 3.5^\circ$ and $y = 0^\circ$ and the other stimulus was in the upper visual hemifield, which required flipping half of the reconstructions across the horizontal meridian.

Quantifying stimulus representations. To quantify the strength of stimulus representations within each reconstruction, we averaged the pixels within each reconstruction located within a $0.9^\circ$ radius disk centered at each stimulus’s known position. This gives us a single value for each stimulus (attended and unattended) on each trial. We then sorted these measurements, which we call “map activation” values because they reflect linear transformations of BOLD activation levels, based on the contrast of the attended and the unattended stimuli (Figs. 5 and 6).

Statistical analysis. For all statistical tests, we used parametric tests (repeated-measures ANOVAs and $t$-tests, where appropriate), followed by 1,000 iterations of a randomized version of the test to derive an empirical null statistic distribution, given our data, from which we compute $P$ values reported throughout the text. If our limited sample satisfies the assumptions of these parametric tests, the $P$ values derived from the empirical null statistic distribution should closely approximate those that would be derived using the parametric test. Especially because of our relatively small sample size, we prefer to rely on the empirical null for recovering $P$ values.

For behavioral analyses (Fig. 2), we computed a two-way repeated-measures ANOVA with attended stimulus contrast and unattended stimulus contrast as factors for each of behavioral accuracy and response time. As a first analysis of BOLD responses, to determine whether it was possible to collapse over sets of trials in which the

---

Fig. 4. Reconstructed spatial maps index stimulus salience and relevance. Image reconstructions are computed for each stimulus separation condition (grouped rows) and each stimulus contrast pair (entries within each $3 \times 3$ group of reconstructions). Within each group of reconstructions, the diagonals (top left to bottom right) are trials with matched contrasts and are useful for visualizing the qualitative effect of behavioral relevance (spatial attention) on map profiles. Even in V1, locations near attended stimuli are represented more strongly than those near unattended stimuli. Going from top to bottom (unattended stimulus) and left to right (attended stimulus) within each group of reconstructions steps through increasing stimulus contrast levels and can be used to infer the sensitivity of each visual field map to visual salience. Color mapping scaled within each region of interest (ROI) independently (see inset, left center). Range subtended by vertical line for each ROI indicates color range (see color bar) for that ROI. BOLD, blood oxygen level dependent; V1–hV4, visual cortex areas; IPS0–IPS3, intraparietal sulcus regions; sPCs, superior precentral sulcus.

J Neurophysiol • doi:10.1152/jn.00059.2018 • www.jn.org

Downloaded from www.physiology.org/journal/jn by $individualUser.givenNames$ $individualUser.surname$ (128.122.115.060) on June 6, 2018.

Copyright © 2018 American Physiological Society. All rights reserved.
irrelevant stimulus contrast varied (see Fig. 6), we computed a two-way repeated-measures ANOVA with attended stimulus contrast and unattended stimulus contrast as factors for each of the stimulus representation activation values for each ROI. We were primarily interested in whether there were any interactions between attended and unattended stimulus contrast, which would have precluded us from collapsing over nonsorted stimulus contrasts. For completeness, $P$ values from our shuffling procedure for both main effects and the interaction for each ROI are presented in Tables 1 and 2. For a subsequent BOLD analysis testing the effects of salience and relevance on map activation across ROIs, we first conducted a three-way ANOVA with factors of stimulus contrast, stimulus identity (attended vs. unattended), and ROI to identify whether there was a difference in attention-related changes in stimulus reconstructions across ROIs (as indicated by interactions between ROI and any other factor). We then performed a follow-up analysis on each ROI by computing a two-way ANOVA with factors of stimulus contrast and stimulus identity (attended vs. unattended) for each ROI.

Fig. 5. Quantifying sensitivity to visual salience and behavioral relevance across cortex. For each stimulus, we averaged the reconstruction activation across all trials where each stimulus appeared at a given contrast (see also Fig. 6). In visual cortex areas V1, V2, V3, and hV4, reconstructed map activation increased with stimulus salience, regardless of whether the stimulus was attended. In all regions of interest (ROIs) except for intraparietal sulcus regions IPS2 and IPS3 and superior precentral sulcus (sPCS), reconstructed map activation also reflected behavioral relevance such that the attended location was more active than the unattended location. Error bars are SE across participants ($n = 8$). Lines indicate best-fit linear contrast response function, and all data are plotted on a log scale. All $P$ values are available in Table 1. Black symbols indicate significant main effects or interaction (corrected via false discovery rate across all ROIs), and gray symbols indicate trend ($P \leq 0.05$, uncorrected). BOLD, blood oxygen level dependent; attn, attended; unattn, unattended.

Fig. 6. Attended stimulus representation does not depend on distractor stimulus contrast. A: reconstructed map activation for attended stimulus location across all salience combinations. Within each plot, linked points correspond to the contrast of the attended stimulus, and the individual linked points indicate the attended stimulus activation at each unattended stimulus contrast level (20%, 40%, 80%). Error bars are SE across participants. We observed no significant interactions (2-way permuted repeated-measures ANOVA within each region of interest (ROI), factors of attended and unattended stimulus contrast, $P \leq 0.041$, minimum $P$ value V3A), so we collapse across unattended stimulus salience within each attended stimulus salience level (Fig. 5). B: same as A, but for unattended stimulus activation, as sorted by attended stimulus contrast within each unattended stimulus contrast level. Again, we did not observe any significant interactions ($P \geq 0.026$, minimum $P$ value V3A), so these are collapsed (Fig. 5). Error bars are SE. All $P$ values are available in Table 2. Black symbols indicate significant main effects or interactions, corrected for multiple comparisons via false discovery rate across regions of interest. Gray symbols indicate trends, defined as $P \leq 0.05$, uncorrected. BOLD, blood oxygen level dependent; Attn, attended; Unattn, unattended.
Table 1. *P* values for sensitivity to visual salience and behavioral relevance (and their interaction)

<table>
<thead>
<tr>
<th>ROI</th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
<th>V3A</th>
<th>hV4</th>
<th>IPS0</th>
<th>IPS1</th>
<th>IPS2</th>
<th>IPS3</th>
<th>sPCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salience</td>
<td>0.003</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.022</td>
<td>&lt;0.001</td>
<td>0.223</td>
<td>0.394</td>
<td>0.371</td>
<td>0.653</td>
<td>0.717</td>
</tr>
<tr>
<td>Relevance</td>
<td>&lt;0.001</td>
<td>0.002</td>
<td>&lt;0.001</td>
<td>0.003</td>
<td>0.005</td>
<td>0.001</td>
<td>0.011</td>
<td>0.071</td>
<td>0.112</td>
<td>0.038</td>
</tr>
<tr>
<td>Interaction</td>
<td>0.062</td>
<td>0.183</td>
<td>0.117</td>
<td>0.166</td>
<td>0.291</td>
<td>0.755</td>
<td>0.527</td>
<td>0.922</td>
<td>0.204</td>
<td>0.644</td>
</tr>
</tbody>
</table>

3-Way ANOVA

<table>
<thead>
<tr>
<th><strong>Salience</strong></th>
<th><strong>Relevance</strong></th>
<th><strong>ROI</strong></th>
<th><strong>Salience X Relevance</strong></th>
<th><strong>Salience X ROI</strong></th>
<th><strong>Relevance X ROI</strong></th>
<th><strong>Salience X Relevance X ROI</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;0.001</td>
<td>0.001</td>
<td>&lt;0.001</td>
<td>0.348</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.166</td>
</tr>
</tbody>
</table>

On each iteration of our shuffling procedure, we shuffled the data values for each participant individually and recomputed the test statistic of interest. To derive *P* values, we computed the percentage of shuffling iterations in which the “null” test statistic was greater than or equal to the measured test statistic (with intact, unshuffled labels). The random number generator was seeded with a single value for all analyses (derived by asking a colleague for a random string of numbers over instant messenger). For ROI analyses, trials were shuffled identically for each ROI.

When appropriate (Figs. 5 and 6), we controlled for multiple comparisons using the false discovery rate (Benjamini and Yekutieli 2001) across all comparisons within an analysis. All error bars reflect SE, unless indicated otherwise.

Data and code availability. All data (behavioral data and extracted trial-level BOLD signal from each ROI after preprocessing) and stimulus presentation and data analysis code necessary to produce figures supporting findings reported in this article are freely accessible online in an Open Science Framework repository (https://osf.io/svuzt), and code is posted on GitHub (github.com/tommysprague/pr2stim_mFiles).

RESULTS

We measured blood oxygenation level-dependent (BOLD) activation patterns from each independently identified retinotopic region using functional magnetic resonance imaging (fMRI) while participants covertly attended one of two visual stimuli (2 patches of randomly oriented dark and light lines), each presented at one of three luminance contrast levels (20%, 40%, and 80%; Fig. 2, A and B), to identify a brief target stimulus (coherent lines that formed a spiral) at the attended location. We maintained behavioral performance at a constant accuracy level of ~80% (Fig. 2D; 2-way permuted repeated-measures ANOVA, *P* values for main effect of attended contrast, unattended contrast, and interaction: 0.359, 0.096, and 0.853, respectively) so that any activation changes we observed did not reflect differences in task difficulty or engagement across experimental conditions. Additionally, response time did not vary with attended or unattended stimulus contrast, or their interaction (Fig. 2E; *P* = 0.926, 0.143, and 0.705, respectively), and the threshold coherence did not vary with the attended stimulus contrast (Fig. 2F; 1-way permuted repeated-measures ANOVA, main effect of attended contrast, *P* = 0.153).

We used a multivariate fMRI image reconstruction technique (an inverted encoding model, IEM) to visualize spatial maps of the visual scene using activation patterns from several attended stimulus location

Table 2. *P* values for dependence of map activation at one stimulus location on contrast of stimulus at the other location

<table>
<thead>
<tr>
<th>V1</th>
<th>V2</th>
<th>V3</th>
<th>V3A</th>
<th>hV4</th>
<th>IPS0</th>
<th>IPS1</th>
<th>IPS2</th>
<th>IPS3</th>
<th>sPCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attended contrast</td>
<td><em>&lt;0.001</em></td>
<td><em>&lt;0.001</em></td>
<td><em>&lt;0.001</em></td>
<td>0.019</td>
<td>0.007</td>
<td>0.762</td>
<td>0.769</td>
<td>0.707</td>
<td>0.399</td>
</tr>
<tr>
<td>Unattended contrast</td>
<td>0.655</td>
<td>0.762</td>
<td>0.758</td>
<td>0.852</td>
<td>0.562</td>
<td>0.520</td>
<td>0.509</td>
<td>0.576</td>
<td>0.976</td>
</tr>
<tr>
<td>Interaction</td>
<td>0.288</td>
<td>0.421</td>
<td>0.814</td>
<td>0.041</td>
<td>0.328</td>
<td>0.383</td>
<td>0.227</td>
<td>0.322</td>
<td>0.147</td>
</tr>
<tr>
<td>Unattended stimulus location</td>
<td>0.085</td>
<td>0.956</td>
<td>0.409</td>
<td>0.281</td>
<td>0.728</td>
<td>0.415</td>
<td>0.579</td>
<td>0.249</td>
<td>0.244</td>
</tr>
<tr>
<td>Attended contrast</td>
<td><em>&lt;0.001</em></td>
<td><em>0.001</em></td>
<td><em>0.001</em></td>
<td>0.283</td>
<td>0.003</td>
<td>0.187</td>
<td>0.191</td>
<td>0.604</td>
<td>0.416</td>
</tr>
<tr>
<td>Unattended contrast</td>
<td>0.490</td>
<td>0.501</td>
<td>0.562</td>
<td>0.026</td>
<td>0.485</td>
<td>0.291</td>
<td>0.764</td>
<td>0.714</td>
<td>0.684</td>
</tr>
</tbody>
</table>

Data are *P* values for all tests reported in Fig. 6. To test whether the map activation at one stimulus location depended on the contrast of the stimulus presented at the other location, we conducted a set of 2-way repeated-measures ANOVAs. First, we tested whether the attended stimulus location map activation varied as a function of attended stimulus contrast (20%, 40%, 80%), unattended stimulus contrast (20%, 40%, 80%), or their interaction. We then tested whether the unattended stimulus location map activation varied as a function of attended stimulus contrast, unattended stimulus contrast, and their interaction. All *P* values were computed against null tests in which we shuffled condition labels within each participant independently 1,000 times and compared the intact *F* values against those derived from the shuffled distribution. Because we ran 1,000 iterations of this shuffling procedure, the minimum accurate quantifiable *P* value was 0.001. We corrected for multiple comparisons across all tests via false discovery rate (bold). Italics indicate trends, defined as *P* ≤ 0.05 (uncorrected).
cortical regions in visual, parietal, and frontal cortex (Sprague and Serences 2013). First, we estimated the spatial sensitivity profile of each voxel within a region of interest (ROI) using data measured from a separate set of “mapping” scans. We then used the resulting sensitivity profiles across all voxels to reconstruct a map of retinotopic space in visual field coordinates from single-trial activation patterns measured during the covert visual attention task (Fig. 3). The spatial profile of activation within these maps can be used to infer whether a given ROI is sensitive to visual salience (i.e., does the spatial profile scale with contrast?; Fig. 1B) and whether it is sensitive to behavioral relevance (i.e., does the spatial profile scale with attention?; Fig. 1D). In principle, both visual salience and behavioral relevance could independently alter the landscape of responses in each ROI. Although this technique involves, as a first step, estimating the spatial sensitivity profile of each fMRI voxel (its receptive field), it is important to note that the goals of this analysis are categorically different from those of more typical voxel receptive field (or “population” receptive field, pRF) analysis methods (Dumoulin and Wandell 2008). In this study, our goal was to characterize how the activation pattern across all voxels, given an independently estimated spatial encoding model, supports a region-level attentional priority map of the visual scene on each trial. Because an RF profile requires many trials to estimate (all visual field positions must be stimulated within a condition), and because we aimed to understand how large-scale activation patterns support attentional priority maps, we focused all analyses on the IEM-based map reconstructions.

We found that image reconstructions systematically tracked the locations of stimuli in the visual field (Fig. 4). Qualitatively, the reconstructions from primary visual cortex (V1) reflected both stimulus salience and behavioral relevance: the reconstructed map activations at stimulus locations scaled both with increasing contrast and with the behavioral relevance of each stimulus. This pattern can be seen in Fig. 4: along the diagonal, where visual salience is equal between the two stimuli, map locations near the attended stimulus are more strongly active than locations near the unattended stimulus. However, in posterior parietal cortex (IPSO), only locations near the attended location were substantially active, with little activation associated with the irrelevant item’s location. Additionally, map activation in posterior parietal cortex did not scale with stimulus contrast. Importantly, even when the unattended stimulus was much more salient than the attended stimulus, only the attended stimulus location was strongly active. This demonstrates that behavioral relevance dominates activation profiles in this area. Overall, occipital retinotopic ROIs (V1–hV4, V3A) show a qualitative pattern similar to that observed in V1, and parietal and frontal ROIs show a pattern similar to that in IPS0.

To quantify these effects, we extracted the mean activation level in each reconstructed map at the known position of each stimulus and then evaluated the main effect of visual salience, the main effect of behavioral relevance, and their interaction using a repeated-measures two-way ANOVA (with P values computed using a randomization test and corrected for multiple comparisons via the false discovery rate, FDR; see MATERIALS AND METHODS, Statistical analyses). If visual salience and behavioral relevance independently contribute to representations of attentional priority, we would expect to find a main effect of salience (contrast) and/or relevance (attention) on map activation in any given ROI, but no interactions between the two.

Reconstructed map activation increased significantly with visual salience in V1–hV4 (Fig. 5; P ≤ 0.003; see also Fig. 6), with no evidence of sensitivity to salience in parietal or frontal regions. On the other hand, map activation increased significantly with behavioral relevance not only in V1–hV4 but also in V3A and IPS0–1 (P ≤ 0.011), with a trend observed in sPCS (P = 0.038, trend defined at α = 0.05 without correction for multiple comparisons). There was no interaction between salience and relevance in any visual area that we evaluated (P ≥ 0.062, minimum P value for V1). Additionally, a three-way repeated-measures ANOVA with factors for salience, relevance, and ROI established that the influence of salience and relevance on map activation significantly varied across ROIs (interaction between salience and ROI: P < 0.001; interaction between relevance and ROI: P < 0.001; all P values for 2- and 3-way ANOVAs available in Table 1).

We also tested whether the map activation near one stimulus location depended on the contrast of the other stimulus. A two-way repeated-measures ANOVA within each ROI on attended and unattended stimulus map activation, each with factors of attended and unattended contrast, yielded no significant interactions (P ≥ 0.124, minimum P value for sPCS; although trends emerged in V3A, defined as α = 0.05 without correction for multiple comparisons; Fig. 6; all P values available in Table 2). Thus the unattended stimulus contrast did not impact reconstructed map activation near the attended stimulus, and vice versa. This rules out a potential alternative operationalization of salience in our data set: one could be interested in defining stimulus salience as how different the two stimuli are (that is, the difference between their relative contrasts). Under such a definition of salience, we would expect to see an interaction between the attended and unattended stimulus contrasts on reconstructed map activation (map activation depends on both the contrast at that location and the contrast of the other stimulus). For example, if the high-contrast stimulus was considered more salient when paired with a low-contrast stimulus, then we would expect to see reconstructed map activation for the high-contrast stimulus increasing as its paired stimulus’s contrast decreases; we do not see this qualitatively in any ROI (Figs. 4 and 6).

DISCUSSION

We tested the hypothesis that attentional priority is guided by dissociable neural representations of the visual salience and the behavioral relevance of different spatial locations in the visual scene. Although previous work has demonstrated sensitivity to bottom-up stimulus salience and/or top-down behavioral relevance in visual (Bogler et al. 2011, 2013; Buracas and Boynton 2007; Burrows and Moore 2009; de Haas et al. 2014; Gandhi et al. 1999; Gouws et al. 2014; Ithiupiripat et al. 2014a, 2014b, 2017; Johannes et al. 1995; Kastner et al. 1999; Kay et al. 2015; Kim et al. 2007; Klein et al. 2014; Mo et al. 2018; Murray 2008; Pestilli et al. 2011; Poltoratski et al. 2017; Somers et al. 1999; Sprague and Serences 2013; Vo et al. 2017; Yildirim et al. 2018; Zhang et al. 2012), parietal (Balan and Gottlieb 2006; Bisley and Goldberg 2003; Constantinidis and Steinmetz 2005; Gottlieb et al. 1998; Iputa et al. 2006; Klein et al. 2014; Serences and Yantis 2007; Sheremata and Silver
2015; Sprague and Serences 2013), and frontal cortex (Bichot and Schall 1999; Katsuki and Constantinidis 2012; Serences and Yantis 2007; Thompson et al. 1997), it remains unexplored how each of these factors interact with one another when parametrically manipulated, as well as how large-scale priority maps subtended by entire visual retinotopic regions support representations of salience, relevance, and their interaction. In this study, we identified representations of visual salience within reconstructed attentional priority maps measured from early visual regions, where map activation scaled with image-computable visual salience. Additionally, we identified representations of behavioral relevance across visual and posterior parietal cortex, with some regions, such as IPS0, exhibiting activation profiles that only index the location of a relevant object, even when an irrelevant but salient object was simultaneously present in the display.

These results provide evidence for a distributed representation of attentional priority, with each retinotopic region supporting a combination of independent representations of the visual salience and behavioral relevance of scene elements (Fecteau and Munoz 2006; Itti and Koch 2001; Serences and Yantis 2006). Models of attentional priority have implemented layers corresponding to image-computable visual salience and task-driven behavioral relevance, but large-scale spatial maps of such features have only been documented in isolation [i.e., manipulations of relevance given equal image-computable salience (de Haas et al. 2014; Jerde et al. 2012; Kastner et al. 1999; Pestilli et al. 2011; Serences and Yantis 2007; Silver et al. 2005, 2007; Somers et al. 1999; Sprague and Serences 2013) or of salience given equal behavioral relevance (Bogler et al. 2011, 2013; Zhang et al. 2012; Poltoratski et al. 2017; White et al. 2017; Yildirim et al. 2018)] or at a local physiological scale [i.e., measurements of single neurons in isolation during manipulation of item salience and/or relevance (Balan and Gottlieb 2006; Bichot and Schall 1999; Bisley and Goldberg 2003, 2010; Burrows and Moore 2009; Constantinidis and Steinmetz 2005; Gottlieb et al. 1998; Ipata et al. 2006; Mazer and Gallant 2003; Thompson et al. 1997; White et al. 2017)]. In this study, we independently manipulated the salience and relevance of multiple scene elements and reconstructed large-scale spatial maps using all voxels within a retinotopic region. Because spatial attention has nuanced and multifaceted effects on voxels tuned to locations across the visual field (de Haas et al. 2014; Kay et al. 2015; Klein et al. 2014; Shermata and Silver 2015; Sprague and Serences 2013; Vo et al. 2017), it was necessary to apply the population-level IEM analysis technique to measure the joint effect of all task-related modulations on spatial maps supported by entire regions. This allowed us to demonstrate that salience and relevance each independently contribute to the representation of attentional priority and that maps become less sensitive to salience across the visual processing hierarchy (Fig. 5). It would be challenging to infer how large-scale activation patterns across entire regions support representations of salience and relevance using forward-modeling techniques, such as population or voxel RF modeling (Dumoulin and Wandell 2008), because an estimate of an RF requires stimulation across the entire visual display over many trials, as in our mapping task (Fig. 3A). We recovered priority maps given independently estimated encoding models on each trial from activation patterns across all voxels within a region. These priority maps necessarily take into account the encoding properties of all voxels used for the analysis under the assumption that selectivity is stable across conditions (see de Haas et al. 2014; Klein et al. 2014; Kay et al. 2015; Shermata and Silver 2015; Sprague and Serences 2013; Vo et al. 2017). Although this assumption may not be rigidly satisfied, it provides a sufficient first-pass approach for characterizing region-level stimulus representations during cognitive tasks on single trials and is similar to the common practice of selecting voxels on the basis of retinotopic selectivity (e.g., using stimulus localizers or visual field quadrants).

In our stimulus setup, image-computable salience could only be defined on the basis of luminance contrast of each stimulus; no other stimulus features were parametrically varied. Other studies have defined stimulus salience on the basis of sudden stimulus onsets or distinct stimulus features among a field of distractors (“singletons”). In these previous studies, neural activity in macaque lateral intraparietal area and FEF exhibited properties consistent with a salience map: neurons respond to abrupt-onset stimuli, but not to stable features of the environment (Balan and Gottlieb 2006; Gottlieb et al. 1998; Kusunoki et al. 2000), and they respond strongly to singleton items among uniform distractors (Bichot and Schall 1999; Thompson et al. 1997). Although we did not see responses in parietal cortex consistent with an image-computable salience map (Figs. 4 and 5), it remains possible that other stimulus manipulations could result in salience representations in parietal and frontal cortex. Several recent fMRI studies presented participants with orientation or motion singletons and found responses consistent with salience maps in visual cortex, similar to our observations, although neither reported any parietal cortex data (Poltoratski et al. 2017; Yildirim et al. 2018).

Indeed, results from other studies suggest that the type of salience manipulation may have a substantial impact on modulations measured in humans with fMRI. In a study in which local contrast dimming among patterned patches was used to define a salient region, activation in early visual field maps reflected the decrease in contrast (but increase in salience by virtue of the high-contrast background) with a decrease in activation, rather than the increase in activation expected if these regions detect salience (Betz et al. 2013). This conflicts with another study in which briefly presented and masked orientation texture patches induced salience-related responses in early visual cortex (Zhang et al. 2012; but see Bogler et al. 2013). Finally, in our study, we only observed salience-related activation profiles in occipital regions of visual cortex (Figs. 4 and 5; V1–hV4). Future studies should compare how different types of bottom-up salience manipulations alter activation profiles of putative priority maps across visual, parietal, and frontal cortex.

In addition, many other studies examining the interaction between behavioral relevance and visual salience use visual search tasks, in which the locus of spatial attention must explore the visual scene on each trial to identify a target stimulus. For example, an animal or human may be required to report the orientation of a bar presented within a green square among a field of bars, each surrounded by green circles. As a salient distractor, one circle might be red. In such a task, the relevant location cannot be known ahead of time by the subject. We designed our study to manipulate the spatial location of covert attention via an endogenous cue (Bisley and Goldberg 2003; Carrasco 2011), as well as the location of a
distracting stimulus of varying levels of salience. Accordingly, we could sort trials on the basis of the location where attention was instructed, which allowed us to quantify changes in map activation levels according to behavioral relevance. When attention is free to roam during a visual search task, all locations may be sampled during a trial, which would be impossible to pull apart using hemodynamic measures. However, recent work has adapted the IEM technique for use with high-temporal-resolution methods like EEG (Foster et al. 2017; Garcia et al. 2013; Samaha et al. 2016). Future studies may be able to probe the wandering spotlight of attention during these kinds of visual search tasks.

Although our observation of no sensitivity to stimulus salience in parietal and frontal cortex is consistent with the interpretation that these regions act as “relevance maps,” there are several other possibilities worth consideration. First, it might be the case that these regions are sensitive to stimulus contrast outside the range used in this study (20–80%). For example, IPS regions may respond in a graded fashion to contrasts between 5% and 15% but have a flat or “step function” response function if sampled only outside that range. Because we wanted complete control over the locus of visual spatial attention without using spatial cues, which could interfere with the spatial reconstructions of relevance and salience maps, we used above-threshold stimuli that could be quickly and unambiguously localized by participants. Perhaps presenting lower contrasts would result in more graded response profiles in parietal and frontal cortex. However, this explanation still requires that spatial attention gives rise to a substantial offset in either the threshold or baseline of neural contrast response functions in these regions. Such a pronounced shift in either of these parameters acts, in effect, as a “relevance” signal within this stimulus range, since most of the variance across trials is described by changes in the relevant stimulus location rather than the salience of the scene at the stimulus locations. Even so, we emphasize that our results remain consistent with an account whereby contrast sensitivity differs between visual and parietal cortex, which could mask our ability to observe sensitivity to “salience” signals more generally. As described above, future studies using alternative image manipulations to modify salience, such as feature contrast, can further resolve the role of parietal and frontal cortex in representing salience and relevance (Poltoratski et al. 2017; Yildirim et al. 2018; Zhang et al. 2012).

Second, anterior regions could pool over a greater spatial extent of the screen, given larger RFs in these areas, and this pooling may compress the strength of responses to stimuli at higher contrasts (Kay et al. 2013; Mackey et al. 2017). This explanation would only account for the feedforward, stimulus-driven responses and must still allow for the contribution of an additional signal highlighting relevant regions of the screen (Kay and Yeatman 2017). That is, the compressive summation would not apply to putatively top-down signals related to spatial attention, only to bottom-up signals scaling with image contrast. Moreover, because we used a single stimulus size in this experiment, it is difficult to unambiguously resolve whether compressive spatial summation or a more general compressive contrast nonlinearity contributes to the lack of contrast sensitivity observed in these regions.

The relative balance of stimulus salience and behavioral relevance within a region may shift under additional task conditions, such as manipulations of cue validity or reward magnitude (Chelazzi et al. 2014; Kahnt et al. 2014; Klyszekjo et al. 2014; MacLean and Giesbrecht 2015a, 2015b; MacLean et al. 2016), or when participants are cued to attend to different feature values rather than spatial positions. We speculate that when behavioral goals are ill defined, maps that are more sensitive to image-computable salience will primarily determine attentional priority and guide behavior (Egeth and Yantis 1997). In contrast, when an observer is highly focused on a specific behavioral goal, maps with stronger representations of behavioral relevance will dominate the computation of attentional priority.

GRANTS
This work was supported by National Eye Institute (NEI) Grant R01-EY025872 (to J. T. Serences), a James S. McDonnell Foundation Scholar Award (to J. T. Serences), NEI Grant F32-EY028438 (to T. C. Sprague), a Howard Hughes Medical Institute Graduate Student Fellowship (to S. Ithipuripat), a Royal Thai Scholarship from Ministry of Science and Technology, Thailand (to S. Ithipuripat), and a National Science Foundation Graduate Research Fellowship (to V. A. Vo).

DISCLOSURES
No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS
T.C.S., S.I., V.A.V., and J.T.S. conceived and designed research; T.C.S. performed experiments; T.C.S. analyzed data; T.C.S., S.I., V.A.V., and J.T.S. interpreted results of experiments; T.C.S. prepared figures; T.C.S. drafted manuscript; T.C.S., S.I., V.A.V., and J.T.S. edited and revised manuscript; T.C.S., S.I., V.A.V., and J.T.S. approved final version of manuscript.

REFERENCES


