Characterizing the effects of feature salience and top-down attention in the early visual system

Sonia Poltoratski1*, Sam Ling2, Devin McCormack1, Frank Tong1

1Vanderbilt Vision Research Center, Psychology Department, Vanderbilt University, Nashville, TN 37240, USA
2Department of Psychological & Brain Sciences, Center for Computational Neuroscience and Neural Technology, Boston University, Boston, MA 02139, USA

Abbreviated Title: Salience and Attention in the Early Visual System

* Corresponding Author: Sonia Poltoratski, sonia.poltoratski@vanderbilt.edu, Vanderbilt University Department of Psychology, 111 21st Ave South, 422 Wilson Hall, Nashville TN 37240.

Copyright © 2017 by the American Physiological Society.
Abstract

The visual system employs a sophisticated balance of attentional mechanisms: salient stimuli are prioritized for visual processing, yet observers can also ignore such stimuli when their goals require directing attention elsewhere. A powerful determinant of visual salience is local feature contrast: if a local region differs from its immediate surround along one or more feature dimensions, it will appear more salient. Here, we used high-resolution fMRI at 7T to characterize the modulatory effects of bottom-up salience and top-down voluntary attention within multiple sites along the early visual pathway, including visual areas V1-V4 and the lateral geniculate nucleus (LGN). Observers viewed arrays of spatially distributed gratings, where one of the gratings immediately to the left or right of fixation differed from all other items in orientation or motion direction, making it salient. To investigate the effects of directed attention, observers were cued to attend to the grating to the left or right of fixation, which was either salient or non-salient. Results revealed reliable additive effects of top-down attention and stimulus-driven salience throughout visual areas V1-hV4. In comparison, the LGN exhibited significant attentional enhancement but was not reliably modulated by orientation- or motion-defined salience. Our findings indicate that top-down effects of spatial attention can influence visual processing at the earliest possible site along the visual pathway, including the LGN, while the processing of orientation- and motion-driven salience primarily involves feature-selective interactions that take place in early cortical visual areas.
New & Noteworthy: While spatial attention allows for specific, goal-driven enhancement of stimuli, salient items outside of the current focus of attention must also be prioritized. We used 7T fMRI to compare salience and spatial attentional enhancement along the early visual hierarchy. We report additive effects of attention and bottom-up salience in early visual areas, suggesting that salience enhancement is not contingent on the observer’s attentional state.

Keywords: fMRI, lateral geniculate nucleus, primary visual cortex, salience, visual attention
At any given moment, our visual system is presented with far more information than it can process, yet we seldom feel that our visual experience is incomplete or degraded. This reflects a sophisticated balance of attentional mechanisms: the observer can willfully guide attention toward a task-relevant item, but is also predisposed to notice salient stimuli that appear outside of the current focus of attention. Our ability to navigate the dynamic visual environment critically depends on this interplay of top-down guidance of spatial attention and bottom-up processing of visually salient information.

The visual salience of an item or portion of the visual field describes the degree to which it can 'grab' attention; this subjective percept is strongly influenced by the local feature differences that occur within the visual scene. Behavioral studies have demonstrated that local differences in color, orientation, size, motion, or other basic features can be rapidly detected in visual search tasks (Nothdurft, 1993). Accordingly, computational models of visual salience have underscored the importance of an initial computation of local feature differences, although models differ on whether feature-selective interactions at early stages of visual processing sufficiently compute a salience map (Li, 2002; Koene and Zhaoping, 2007; Zhang et al., 2012), or whether a summation of various feature-contrast maps occurs at a higher stage of the visual pathway (Itti and Koch, 2001a). These models can predict where people are more likely to look when viewing natural scenes (Itti and Koch, 2001a), though cognitive goals and top-down factors can also exert powerful influence (Henderson, 2003).

Neurophysiological studies have found that early visual areas are strongly modulated by local feature contrast. The response of a V1 neuron, for example, is suppressed when a presented stimulus extends beyond the neuron’s receptive field and into the surround; this suppression is orientation-tuned, such that greater suppression occurs when the orientation of the surround matches that of the center than when the center and surround orientations are orthogonal (Blakemore and Tobin, 1972; Nelson...
and Frost, 1978; Cavanaugh, 2002). This form of feature-tuned surround suppression has been shown to emerge for a variety of visual features (Blakemore and Tobin, 1972; Allman et al., 1985; Chao-Yi and Wu, 1994; DeAngelis et al., 1994; Kapadia et al., 1995; Zipser et al., 1996; Jones et al., 2001). Multiple mechanisms likely contribute to feature-tuned surround suppression, including shorter-range interactions arising from horizontal connections between feature-selective neurons within V1 as well as more distal interactions that rely on feedback from higher extrastriate areas (Gilbert and Wiesel, 1979; Lamme, 1995; Cavanaugh et al., 2002; Bair et al., 2003; Angelucci and Bressloff, 2006a). The computation of local feature contrast can be explained in terms of greater mutual inhibition, or divisive normalization (Carandini and Heeger, 2011), among neurons that share similar feature preferences. Thus, a local region that differs in feature content from its immediate surround should evoke a stronger visual response due to a release from feature-tuned surround suppression.

Human neuroimaging studies have investigated the effects of orientation-selective surround suppression and orientation-defined salience, using both simple and more complex displays. Studies employing large target stimuli with immediately abutting surrounds have reported powerful effects of surround suppression (Zenger-Landolt and Heeger, 2003) as well as a more modest feature-selective component of surround suppression (e.g. McDonald et al., 2009). Other studies have tested for effects of orientation-defined salience using more complex displays of multiple spatially separated gratings or lines, akin to the displays commonly used in behavioral investigations of attention and visual search (e.g. Nothdurft, 1993). Results from these neuroimaging studies have been mixed: some studies find that a salient, uniquely oriented item evokes stronger responses in V1 (e.g., Zhang et al., 2012; Schallmo et al., 2016), while others find no reliable differences in early visual areas (Beck and Kastner, 2005; Bogler et al., 2013) or more complex interactions that depend on top-down spatial attention (Hopf et
SALIENCE AND ATTENTION IN THE EARLY VISUAL SYSTEM

al., 2004; Flevaris and Murray, 2015). Our understanding of visual salience and its
neural bases relies critically on testing with multi-item displays, which inform much of our
knowledge of the mechanisms of attention and visual search. One factor to consider is
that feature-tuned suppression may be more difficult to detect with widely separated
items, as surround suppression effects in V1 are known to decrease as a function of
retinotopic distance (Bair et al., 2003; Adesnik et al., 2012; Shushruth et al., 2013).

Another consideration is that some prior studies have employed more complex visual
tasks, raising the possibility that visual attention may interact with the processing of
salient visual information.

Prioritization of visual stimuli is determined not only by stimulus-driven factors but
also by voluntary selective attention, by which one can enhance the processing of stimuli
corresponding to his or her specific goals and needs. While frontal and parietal networks
have been implicated in the control of attention (for reviews: Kastner and Ungerleider,
2000; Corbetta and Shulman, 2002; Silver and Kastner, 2009; Squire et al., 2013),
considerable work has shown that attention strongly modulates responses in early
cortical visual areas, and even the lateral geniculate nucleus (LGN; Brefczynski and
DeYoe, 1999; Gandhi et al., 1999; O'Connor et al., 2002; Schneider and Kastner, 2009;
Ling et al., 2015).

Our goal in this fMRI study was to determine how directed spatial attention and
feature-defined salience modulate responses at early stages of visual processing. One
possibility is that once an item is attended, its salience does not provide any further
enhancement. Alternatively, it could be that the effects of top-down attention and
feature-based salience are simply summed to determine the overall prioritization of a
stimulus in the visual field. We also sought to determine the earliest stages at which
salience and top-down attention would lead to reliable modulations of visual activity.
Recent work from our lab has demonstrated orientation-selective responses in the
human LGN as well as modulatory effects of attention (Ling et al., 2015).

Neurophysiological studies in animals suggest the presence of coarse selectivity in LGN (Smith et al., 1990; Xu et al., 2002; Suematsu et al., 2012; Cheong et al., 2013), arising primarily from elongation of the RF (Leventhal and Schall, 1983). Moreover, there is some evidence to suggest that orientation-tuned surround suppression emerges within cat LGN (Jones et al., 2000; Naito et al., 2007). Therefore, it seemed possible that the LGN might be sensitive to orientation-defined salience. However, feature-selective interactions among spatially distributed items require either long-range connections between feature-tuned neurons or feedback from higher visual regions (Nassi et al., 2013; Nurminen and Angelucci, 2014); these effects may be restricted to cortical visual areas, which have larger receptive fields and effective surrounds.

We used high-resolution fMRI at 7 Tesla to quantify the effects of top-down spatial attention and feature-defined salience at multiple levels of the visual pathway, including areas V1 through hV4 and the lateral geniculate nucleus (LGN). Observers viewed multi-item displays that contained a single salient grating with a unique orientation or motion direction, and were cued to attend to either the salient or a non-salient grating that was located immediately to the left or right of central fixation. The spatial separation of these elements emphasized longer-range feature-tuned contextual interactions. By testing both orientation- and motion-defined salience, we could evaluate the generality of the effects of feature-defined salience, as most previous research has have focused exclusively on orientation processing (Hopf et al., 2004; McDonald et al., 2009; Zhang et al., 2012; Bogler et al., 2013; Flevaris and Murray, 2015; Schallmo et al., 2016); but see also (Harrison et al., 2007).

Materials & Methods

Participants

Six healthy adults (ages 22 to 31, one female) participated in Experiment 1, and
six (ages 22 to 33, three females) participated in Experiment 2. Three subjects participated in both experiments. All participants had normal or corrected-to-normal vision, and were compensated for their role. All aspects of the study were approved by the Vanderbilt University Institutional Review Board.

Stimuli & Task: Experiment 1

Our experimental displays, illustrated in Figure 1A, were created using Matlab and Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Displays consisted of a 3 x 4 array of Gabor gratings presented at 100% Michelson contrast, 1.5 cycles/degree, and Gaussian envelope $\sigma=0.4^\circ$. The stimuli measured $3.8^\circ$ center-to-center. If one considers the effectively visible portions of the Gabor to span 5 standard deviations in width, where the contrast at the edge would dip to 4%, then each Gabor grating can be considered to have an effective width of $1^\circ$ with a gap of $1.8^\circ$ between the Gabors.

In each block, all gratings except for one appeared with a common orientation, either vertical or horizontal; the other, salient, grating had an orthogonal orientation. The salient grating appeared at one of two possible target locations, immediately to the left or right of the fixation point ($0.5^\circ$ in diameter with a central dot of $0.175^\circ$) – we refer to these as the left and right ‘target locations’ for our region of interest (ROI) analyses. All patches flickered on/off in 200ms intervals (i.e., 2.5 Hz), with the spatial phase of the Gabor randomized on each presentation.

In each experimental block, the participant was shown a cue at fixation, instructing them to covertly attend to one of the target gratings while keeping their eyes focused on the central fixation point. The attentional cue consisted of a pair of dots ($0.1^\circ$) that appeared to the left and right of fixation. One of these dots was black, and the other white; each participant was told to attend to the side marked by one of these two colors throughout the experiment. The spatially balanced design of this fixation cue ensured equivalent stimulus-driven activity in each hemifield and avoided potential effects of
exogenous cuing, which could occur with a single lateralized cue (Jehee and Tong, 2011; Jehee et al., 2012). The attended cue color was counterbalanced across participants.

Participants were asked to perform an attentionally engaging contrast-decrement detection task on the cued stimulus. We chose this task to direct covert spatial attention to the cued grating, and to minimize the potential influences of feature-based attention. In previous work, we have shown that the performance of an attentional task on the orientation of a grating leads to a strong enhancement of orientation-selective responses in V1-V4, while tasks that require attending to the contrast of a grating do not (Jehee and Tong, 2011). Contrast decrements occurred at independent intervals at each of the two target locations throughout the experimental block, but participants were instructed to respond to decrements only at the cued location via a button box. The decrement occurred 8 times per 16s block at randomly determined intervals, lasting for the full 200ms duration of that ‘on’ interval with a minimum 800ms time difference between targets.

**Stimuli & Task: Experiment 2**

The stimulus parameters were similar to those of Experiment 1, except for two key differences: motion direction, rather than orientation, was used to define salience, and the drifting Gabor patches were continuously presented at 30% contrast (see Figure 1B). All of the patches were oriented horizontally and phase-randomized, and the gratings drifted either upward or downward at a speed of 5 degrees/second (temporal frequency, 7.5 Hz) within a stationary window. The salient patch moved in the opposite direction relative to the motion of all other Gabor gratings in the array. We chose upward-downward motion directions to minimize the likelihood of inducing involuntary optokinetic nystagmus (OKN; Honrubia et al., 1968), and to further ensure that if a small
eye movement was occasionally induced, it would not be directed toward either of the lateralized target gratings. Participants were asked to perform an attentionally engaging contrast-decrement detection task on the cued stimulus. The contrast decrement lasted for 400ms, and occurred on 20% of these implicit 400ms intervals within each 16s stimulus block.

In a separate behavioral eye-tracking session, we confirmed that our participants could maintain stable fixation while viewing these arrays of moving gratings. Participants performed the same experimental task on displays that matched the stimuli and timing (16s blocks) of the fMRI experiment. Overall, our subjects kept their eyes fixed well within the 0.5° fixation dot. Horizontal eye movement deviations (reported as the standard deviation from the true fixation position), ranged from 0.07-0.33° across the 6 participants (median 0.15°); vertical deviations ranged from 0.11-0.48° (median 0.23°).

**Experimental Design and Procedure**

Apart from differences in the stimuli, the design of both experiments was identical. There were three experimental factors that consisted of: 1) location of the salient grating (left or right of fixation), 2) location to be spatially attended (left or right grating), and 3) the context feature (horizontal or vertical orientation in Experiment 1, upward or downward motion in Experiment 2). This 2 x 2 x 2 factorial design led to 8 experimental conditions, which were presented in a randomized order within each run. Each experimental run followed a 16s block design (272s duration), with each of the 8 experimental conditions occurring once, interleaved with 16s fixation-rest periods that also occurred at the beginning and end of each run. A fixation circle remained present throughout the experimental run; the spatial attention cue appeared 1s prior to the onset of each stimulus block, informing participants to direct their attention towards the left or right grating. Participants were informed of their performance accuracy at the end of
each run, and the magnitude of the contrast decrement was adjusted between runs so that the detection hit rate for each subject fell in a range from approximately 70-90%; across subjects, the magnitude of the contrast decrement ranged from 25% to 40% of the original contrast value.

*Behavioral Performance*

In Experiment 1, performance at the contrast detection task led to a mean hit rate across individual subjects of 76.4% (st. dev. = 12.1%). Contrast decrements applied to the salient target in the display were detected with 77.4% accuracy, while those applied to non-salient targets yielded 75.4% correct performance. There was no significant difference in performance for salient and non-salient targets across the group ($t(5) = 0.81, p = 0.45$). In Experiment 2, subjects averaged 83.4% hit rate (st. dev. = 7.1%), 83.7% correct on salient targets and 82.5% correct on non-salient targets; again, there was no reliable difference in performance for the salient and non-salient targets ($t(5) = .59, p = 0.58$). Our task, which requires contrast discrimination using supra-threshold contrast changes, was designed to manipulate the locus of spatial attention rather than to assess sensitivity to contrast, which has been shown to be enhanced at salient locations. (e.g. Kapadia et al., 1995; Joo et al., 2012). These behavioral results suggest that participants were able attend selectively to the non-salient grating location when it was the target, and were not distracted by the presence of the salient grating.

*fMRI Data Acquisition*

Data were acquired using a Philips Achieva 7-Tesla MRI scanner at the Vanderbilt University Institute of Imaging Science (VUIIS), with a 32-channel head coil equipped for SENSE imaging. The functional scans employed single-shot gradient-echo echoplanar imaging to measure BOLD activity, and were aligned approximately parallel to the AC-PC line to best capture the LGN and occipital pole. Subjects were scanned
using either 1.5 mm isotropic voxel resolution in 20 slices (6 subjects in Experiment 1, 4 in Experiment 2) or a 2 mm × 1.875 mm × 1.875 mm voxel resolution in 36 slices (2 subjects in Experiment 2). The 1.5 mm thickness scans used the following parameters: 2 s TR, 26 ms TE, 75° flip angle, 192 mm FOV, with no gap. The 2 mm scans used identical parameters, except that the FOV was 210 mm. Twelve to 16 runs of functional data and 2-3 localizer runs were collected for each subject. The spatial extent of our slice prescription allowed us to monitor BOLD activity in the LGN and areas V1, V2, V3 and hV4. Area V3A often appeared outside of this slice prescription and was therefore not included in this study.

Functional ROI Definition

Cortical visual areas V1-hV4 were demarcated using standard retinotopy procedures, using data acquired from separate scan sessions at 3-Tesla (Engel et al., 1997; Wandell et al., 2007). We used a typical phase-encoded design in which subjects fixated while they viewed flickering checkerboards consisting of rotating wedges to map polar angle and expanding rings to map eccentricity. Retinotopy data was acquired using a Philips 3T Intera Achieva MRI scanner equipped with an 8-channel coil. Subjects were scanned using 3 mm isotropic resolution (TR 2 s, TE 35 ms, flip angle 80°, 28 slices, 192 x 192 FOV). Boundaries between retinotopic areas V1-hV4 were delineated by hand, by identifying reversals in the phase of the polar angle map measurements; the resulting ROIs were aligned to the functional space of the current experiment using FSL and Freesurfer software.

In the experimental scan session, we ran 2-3 runs of a visual localizer to identify the target regions of interest corresponding to the spatial extent of the left and right gratings. This involved presenting flickering checkerboards at full contrast within a Gaussian contrast-envelope at each of the two target locations, with alternating 16-s
cycles of left or right stimulation, as well as a 16 s fixation period at the beginning and end of each run. Individual checks were 0.75° in width, and contrast-reversed at a rate of 5Hz. Subjects did not perform a task during localizer runs, and were instructed to keep their eyes on a central fixation point. Cortical ROIs were selected from the conjunction of retinotopy and a statistical map of the left vs. right contrast of our functional localizer. We report results from the 100 most functionally selective voxels as defined by the t-statistic map in each lateralized ROI in early visual areas.

The lateral geniculate nucleus was defined functionally from the same localizer contrast, using a t-value threshold of no less than 2.8; thresholds for each subject were selected to yield distinct, continuous clusters of voxels such that left and right hemisphere nuclei were generally aligned dorsally and were maximally lateralized, so as to avoid the inclusion of the pulvinar region. As the LGN cannot be localized anatomically from T1- or T2-weighted images, we cannot be fully sure that our regions of interest do not include other portions of the thalamus. However, the LGN is more readily activated by visual stimulation than other subcortical regions, and there is evidence to suggest that functional localizers that rely on passive viewing, as ours did, do not activate pulvinar activity as strongly as the LGN (Kastner et al., 2004).

In sessions using a 1.5 mm isotropic voxel size, the bilateral LGN region of interest consisted of an average of 72.9 voxels (stdev. = 16.8); in the two participants who were tested using a larger voxel size of 2 mm x 1.875 mm x 1.875 mm voxel size, the LGN region of interest encompassed an average of 42.5 voxels.

fMRI Analysis: Preprocessing

Data were preprocessed using FSL and Freesurfer tools (documented and freely available for download at http://surfer.nmr.mgh.harvard.edu), beginning with 3D motion correction and linear trend removal, followed by a high-pass filter cutoff of 60s. Functional images were registered to a reconstructed anatomical space for each subject;
this registration was first automated in FSL and then checked and corrected by hand.

This allowed the alignment of the current fMRI data to the retinotopy data, which was collected in a separate session. The functional localizer was spatially smoothed using a 1-mm Gaussian kernel; no spatial smoothing was done for the experimental runs.

Further analyses were conducted using a custom Matlab processing stream. For each ROI-based analysis, data were masked using the labels defined from the retinotopy and functional localizer, which corresponded to the left and right target grating locations in each visual area. Each voxel’s intensities were normalized by the mean of the time series, converting to mean percent signal change within each run. Outliers were defined as time points for which the voxel’s response measured more than 3 times its standard deviation from its mean, and were Winsorised (Hastings et al., 1947). This condition-blind preprocessing step minimizes the impact of rare spikes in MR intensity while preserving the temporal structure of the responses in each voxel. Only a small fraction of a percentage of data points in the current experiments were marked as outliers (0.26-0.41% across Experiments and ROIs). Further, voxels that left the volume at any time point due to head motion were excluded from that run’s analysis. Additionally, we found that during one session in Experiment 1, a combination of slice placement and head motion caused some regions to intermittently clip the edge of the volume. For this subject, we excluded 2 runs from the analysis of bilateral V2 and right hV4, based on drop-out of visually selective responses that was restricted to these ROIs.

**fMRI Analysis: Mean BOLD**

To calculate the average mean BOLD response for each experimental condition, we first converted the MR time series of each voxel from signal intensity units to units of percent signal change. Next, we calculated the average BOLD amplitude for each block, after shifting the response period by 2 TRs to account for hemodynamic lag. To account
for baseline differences preceding each condition’s onset, the average of the 2 TRs immediately preceding each block was subtracted from its mean. For every stimulus block, we obtained a measure of the mean BOLD response in each ROI, and further calculated the overall BOLD response across all blocks for each experimental condition.

Conditions were defined relative to lateralized ROI responses such that, for example, a block in which the left target patch was salient was labeled ‘salient’ in the right hemisphere and ‘non-salient’ in the left hemisphere; data were pooled in this way across the left and right hemisphere ROIs. We also combined data across specific feature values (e.g. when the salient target was vertical vs. horizontal in Experiment 1), which were not pertinent to our hypotheses. In each bilateral region, we performed a 2 x 2 repeated measures ANOVA to characterize the effects of salience and spatial attention, as well as their interaction.

Results

fMRI: Experiment 1

In this experiment, orientation was used to define one salient target location in an array of gratings, while the participant spatially attended to the salient target or to a non-salient target in the contralateral hemifield (see Figure 1A). We compared fMRI responses in early visual areas evoked by the salient grating and by the contralateral non-salient grating across attended and unattended conditions, pooling the data across the left and right ROIs.

The time course of mean BOLD responses in the LGN and areas V1-hV4 are shown in the top panel of Figure 2B, while the mean response amplitudes observed in each block are plotted in the bottom panel. Figure 2A shows the modulatory effects of both attention and salience as average difference in BOLD response. Spatial attention led to positive increases in the BOLD response throughout the visual pathway, in both subcortical and cortical regions of interest. We observed reliable effects of top-down
attentional modulation in all individual cortical visual areas (V1: $F(1,5) = 7.59, p < 0.041$; V2: $F(1,5) = 8.08, p < 0.037$; V3: $F(1,5) = 23.7, p < 0.005$; hV4: $F(1,5) = 19.2, p < 0.008$). This effect was likewise present in the LGN ($F(1,5) = 9.11, p < 0.030$), consistent with previous fMRI reports that the human LGN can be reliably modulated by spatial attention (O’Connor et al., 2002; Schneider and Kastner, 2009; Ling et al., 2015).

In contrast, we found that the effect of salience was evident only in the visual cortex. No reliable difference between salient and non-salient items was observed in the LGN ($F(1,5) = 0.028, p = 0.87$). Early visual cortical areas, however, exhibited clear and reliable effects of salience, with higher mean BOLD responses to the salient grating than to the non-salient grating that matched the orientation of the surrounding context (V1: $F(1,5) = 9.47, p < 0.003$; V2: $F(1,5) = 8.53, p < 0.034$; V3: $F(1,5) = 42.8, p < 0.002$; hV4: $F(1,5) = 7.35, p < 0.043$).

Interestingly, the main effects of salience and spatial attention did not significantly interact in any of our measured ROIs (V1: $F(1,5) = 1.83, p = 0.24$; V2: $F(1,5) = 0.14, p = 0.73$; V3: $F(1,5) = 2.34, p = 0.19$; hV4: $F(1,5) = 0.12, p = 0.74$). That is, the enhancement of salient items is not contingent on the observer’s attentional state or goals. The LGN, which did not show a significant main effect of salience but was reliably modulated by attention, likewise showed no interaction effect ($F(1,5) = 1.07, p = 0.35$).

As shown in Figure 2A, comparison of the magnitude of these effects across cortical visual areas reveals a significant difference between regions (ANOVA, $F(3,15) = 5.74; p < .009$), suggesting a trend of increasing attention modulation along the visual hierarchy (F-test, $t(2) = 9.00, p < .013$). There appeared to be no difference in the magnitude of salience modulation across visual areas V1 through V4 ($F(3,15) = 1.68, p = 0.21$). Including the LGN in this comparison, however, did yield significant differences across regions of interest ($F(4,29) = 7.8, p < 8.0 \times 10^{-4}$). Specifically, the salience modulation of the LGN was not significantly different from zero ($t(5) = 0.16, p = 0.87$),
and significantly weaker than the salience effect observed in V1 ($t(5) = 4.7, p < 0.0054$).

We performed a Bayes factor analysis to estimate the likelihood that the results from the LGN could have arisen from a null effect of salience. We calculated the JZS Bayes Factor (Rouder et al., 2009) applying a scale factor of 1 for the prior on effect size for the alternative hypothesis, and obtained a value of 3.44 in favor of the null hypothesis. It has been suggested that odds factors greater than 3 should be considered as evidence in favor of a hypothesis (Jeffreys, 1961); thus, the LGN data are not strongly conclusive but do tend to favor the null hypothesis.

To further test for the possible presence of an interaction effect between salience and spatial attention, we performed an ANOVA on the data of individual subjects, using the fMRI response amplitudes observed on individual fMRI blocks for each participant (Figure 6). Consistent with our group-analysis results, we observed statistically significant main effects of salience and top-down attention in early visual areas for the majority of individual subjects. However, the interaction between salience and attention did not reach statistical significance in any ROI for any subject. To illustrate the quality of data collected, as well as the differences between measurements of the LGN and of cortex, representative single-subject data from this Experiment is shown in Figure 3A/B.

All preceding analyses were performed on the mean activations of the functional ROIs; in the cortex, these were selected from individual retinotopic regions, and defined as the 100 most selective voxels in each hemisphere based on independent localizer runs. We assessed whether this ROI size criteria had a meaningful effect on the pattern of our results by calculating the magnitude of attentional and salience-based modulation for a wide range of ROI sizes. Figure 4 plots the magnitude of attention and salience-based enhancement as a function of the number of voxels selected from each ROI. As can be seen, these effects are broadly consistent across a wide range of ROI sizes.

The results of Experiment 1 suggest that effects of orientation-defined salience
emerge in the primary visual cortex, with a similar degree of enhancement observed in higher extrastriate visual areas. Although orientation-specific responses have recently been demonstrated in the human LGN (Ling et al., 2015), here we find no evidence of enhanced responses to orientation-defined salience at this subcortical site. However, we do find that spatial attention reliably modulates responses in the LGN, in agreement with previous fMRI studies (O’Conner et al., 2002; Ling et al., 2015). The finding that modulatory effects of visual salience and top-down attention emerged at different levels of the visual pathway provides support for the proposal that these mechanisms operate independently and are functionally distinct. Consistent with this proposal, the modulatory influences of spatial attention and orientation-defined salience appeared to be separable and additive in early visual areas of interest.

fMRI: Experiment 2

Do the effects of orientation-based salience generalize to other feature domains? Most studies that find salience enhancement in early visual areas have focused on orientation as the defining feature (Li, 2002; Joo et al., 2012; Zhang et al., 2012; Schallmo et al., 2016); however, direction-selective contextual interactions in V1 have been reported for spatially separated moving stimuli (Harrison et al., 2007). In Experiment 2, we were motivated to test whether motion-defined salience would also yield similar effects of top-down attention and stimulus-driven salience across the visual hierarchy. In this experiment, one Gabor grating drifted in a direction opposite to that of all other gratings in the display and was deemed salient. Meanwhile, observers performed an attentionally demanding task on either the salient item or on a non-salient item, as illustrated in Figure 1B.

As can be seen in Figure 5, mean BOLD responses were consistently greater for attended than unattended items. A repeated-measures ANOVA indicated that this
modulatory effect of attention was statistically significant in all regions of interest, including the lateral geniculate nucleus (LGN: $F(1,5) = 27.5, p < .004$; V1: $F(1,5) = 47.9$, $p < 9.7 \times 10^{-4}$; V2: $F(1,5) = 44.0, p < 0.0013$; V3: $F(1,5) = 37.6, p < 0.002$; hV4: $F(1,5) = 96.6, p < 1.9 \times 10^{-4}$). While absolute BOLD amplitudes in the LGN were variable across subjects, as evidenced by the size of the error bars in Figure 5B (which correspond to +/- 1 S.E.M. across the six subjects), the within-subject effect of attention was statistically reliable in the LGN. Moreover, the effect of attention appeared highly consistent in both LGN and visual cortex, with every participant showing attentional modulations in the predicted direction. This is illustrated in Figure 5A, which depicts individual subjects’ effect amplitudes as grey dots overlaid on the mean effect across ROIs; representative individual subject data for Experiment 2 is shown in Figure 3C/D. In the visual cortex, we again observed significant modulation by salience: non-salient gratings that drifted in the same direction as the contextual gratings elicited weaker mean BOLD responses than did items that drifted in the opposite direction. Significant enhancement of salient items was observed in each of the cortical regions of interest, beginning in V1 ($F(1,5) = 33.7, p < .003$) and persisting through V2-hV4 (V2: $F(1,5) = 10.0, p < 0.026$; V3: $F(1,5) = 42.5, p < 0.002$; hV4: $F(1,5) = 18.1, p < .009$). However, this motion-defined salience did not significantly modulate activity in the LGN ($F(1,5) = 0.14, p = 0.72$). Critically, the enhancements of BOLD activity elicited by attention and salience did not appear to interact in this experiment. Analysis of variance indicated that salience and attention conditions did not lead to a significant interaction effect in any of the cortical regions of interest (V1: $F(1,5) = .015, p = 0.91$; V2: $F(1,5) = 4.4, p = 0.09$; V3: $F(1,5) = 0.083, p = 0.78$; hV4: $F(1,5) = .62, p = 0.47$), nor in the LGN $F(1,5) = 1.90, p = 0.23$). Thus, the degree of enhancement observed for salient items appeared comparable under conditions of attention and inattention.
Figure 5A shows the modulatory effects of spatial attention and of salience in Experiment 2 for each region of interest. Similar to the results of the first experiment, we see a trend toward increasing attentional modulation as one ascends from V1 to V4 (ANOVA $F(3,15) = 4.26, p = 0.023$; F-test $t(2) = 7.19, p < .019$). Salience modulations appear similar in magnitude across areas V1-hV4, and no reliable difference was found among these cortical ROIs ($F(3,15) = .38, p = 0.77$). A difference emerged across brain areas when data from the LGN was included in the analysis of variance ($F(4,29) = 7.46, p < 9.0 \times 10^{-4}$) and paired comparisons indicated that salience modulation was significantly weaker in the LGN than in V1 ($t(5) = 4.88, p < .0046$). As in the previous experiment, salience modulation in the LGN did not significantly differ from zero ($t(5) = .38, p = 0.72$). In this case, a Bayes factor analysis indicated a value of 3.25, in moderate favor of the null hypothesis. We again note the difficulty in interpreting a null effect in the LGN: small effect sizes, when present, are difficult to detect, and the increased physiological noise when measuring subcortical activity can impede the reliability of fMRI measures. However, the magnitude of salience modulation appears clustered around zero for the majority (5/6) of our individual subjects (Figure 5A), in contrast with individual effects of attention in the LGN or of salience in cortex.

Overall, the two experiments yielded very similar patterns of results, demonstrating that both orientation- and motion-defined salience lead to common effects of enhancement in the early visual cortex. These effects did not significantly interact in any region of interest, neither in the group analysis nor in analysis results for any individual subject, suggesting that stimulus-driven salience and top-down attention provide independent sources of modulatory influence at early stages of the visual system.

Discussion
In two experiments, we used high-field fMRI to characterize the effects of top-down attention and feature-defined salience at multiple levels of the visual hierarchy. We found that orientation- and motion-defined salience consistently enhanced responses to stimuli in areas V1 through hV4, implying a common mechanism for detecting local feature differences across spatially disparate stimuli. These feature-tuned contextual interactions are critical for the computation of the local salience of a region, as predicted by modeling work (Itti and Koch, 2001b; Li, 2002). Additionally, we found that directed spatial attention enhanced responses in all regions of interest, including the LGN; our findings add to a growing number of studies indicating that activity in the LGN can be altered by top-down attentional goals (O’Connor et al., 2002; McAlonan et al., 2008; Schneider and Kastner, 2009; Ling et al., 2015). In contrast, we did not find evidence that LGN activity is reliably modulated by feature-defined salience. Although one must be cautious about the interpretation of null effects, the absence of a statistically reliable effect of salience in the LGN is consistent with the proposal that long-range feature-selective interactions depend on cortical mechanisms (Bair et al., 2003; Shushruth et al., 2009; Nassi et al., 2013; Shushruth et al., 2013). Our results support the proposal that at the earliest stages of visual processing, salience is distinct from mechanisms of top-down attention. Consistent with this view, we found that the effects of salience did not lead to significant interactions with the effects of attention in any of our regions of interest. Our findings suggest that the effects of top-down attention and bottom-up salience are summed in a simple additive manner, such that both mechanisms distinctly inform the prioritization of items within a visual scene. These results concur with a recent fMRI study by Schallmo et al. (Schallmo et al., 2016) though it should be noted that more complex interactions between bottom-up mechanisms and top-down attention may occur if attention is allowed to spread to other neighboring items. For example, feature-based...
attention may counteract the influence of feature-tuned suppression with certain stimulus configurations and tasks. Flevaris and Murray (2015) found that an attended target grating evoked greater responses when flanked by orthogonally oriented gratings than when the target and flankers were iso-oriented. When participants attended to one of the flanking gratings, however, attention appeared to spread to the target if the gratings shared a common orientation. This reported interaction presumably emerged due to the spatial spreading of feature-based attention (Saenz et al., 2002). In the present study, attention was cued to shift between target locations in separate hemifields, thereby minimizing the spread of attention between the lateralized target regions. Under these conditions, we observed separate additive effects of top-down attention and salience at each of the target locations. We should note, however, that the limited range of contrast levels in the current experiments may bias us toward finding additive effects if the neural response function is saturated (Carrasco, 2006).

Our results support the view that processing of feature-defined salience in V1 relies on horizontal interactions within the region (Stettler et al., 2002; Adesnik et al., 2012) as well as more long-range effects of feedback from higher extrastriate areas to V1 (Bair et al., 2003; Angelucci and Bressloff, 2006b; Nassi et al., 2013). Such feature-selective interactions are believed to underlie a variety of early contextual effects (Stettler et al., 2002) including tilt repulsion (Dragoi et al., 2001; Jin et al., 2005), collinear enhancement (Kapadia et al., 1995), and figure-ground processing (Lamme, 1995); the current experiments considered how these early visual interactions may function over spatially disparate elements in a scene, which is critical toward building a behaviorally relevant representation of salience. Of course, further processing of salience occurs in higher-level extrastriate and parietal regions as the observer orients his or her attention during visual search (Melloni et al., 2012), integrates many features in a naturalistic environment (Bogler et al., 2011), shifts attention (Kincade et al., 2005), or
plans eye movements (Gottlieb et al., 1998; Mazer and Gallant, 2003; Fectueau and
Munoz, 2006). The current work informs the wider study of salience by investigating the
early cortical stages involved in these processes, and by demonstrating that the
representation of this information is unaffected by the observer's attentional focus.

Our results in the LGN suggest that feature-selective surround effects in this
region are either negligible or too weak to be reliably detected in the current paradigm,
which relied on spatially separated gratings to manipulate salience. Existing
neurophysiological work is consistent with these views. Studies of the LGN in cats and
monkeys have demonstrated size-tuned responses to gratings due to strong
suppression from the immediate surround, which likely arises from mechanisms
originating from the retina (Alitto and Usrey, 2008) as well as effects of cortical feedback
(Jones et al., 2012). While some orientation bias has been reported in LGN neurons in
several species (e.g. cat: Suematsu et al., 2012; owl monkey: Xu et al., 2002; Cheong et
al., 2013; and macaque: Smith et al., 1990), positive effects of feature-tuned
suppression in this structure have been predominantly reported in the cat (Cudeiro and
Sillito, 1996; Jones et al., 2000; Naito et al., 2007). Given the known anatomical
differences between cat and primate LGN, our null effect could reflect a lack of feature-
tuned surround suppression in the region. It is also possible that feature-tuned
suppression in the LGN may operate at a more local spatial scale than our Gabor array
displays were designed to probe (Angelucci and Bressloff, 2006b), as suggested by our
group’s positive report of modulatory effects of orientation masking in this structure (Ling
et al., 2015). Alternatively, it may be that particular elements of our experimental design
precluded us from detecting modulatory effects: we were limited in testing only one
contrast level in each experiment, and using a constant spatial frequency (1.5cpd)
throughout the study.
These findings contribute to a broader understanding of how surround suppression functions toward a behaviorally relevant representation of the visual scene. The effects of surround suppression have been shown to fall off with distance; however, modulations in apparent contrast of a central stimulus can be detected even with spatial separation of several degrees between the center and surround (Cannon and Fullenkamp, 1991; Petrov and McKee, 2006). While the generalization of surround suppression to natural image or movie inputs has posed a challenge in vision research, recent work has suggested that the suppression may be gated by the degree of redundancy in a natural image (Coen-Cagli et al., 2015). Homogeneity in a natural image was found to elicit stronger suppression, which may act to reduce redundancy in visual input and code natural input more efficiently (Vinje and Gallant, 2000). fMRI studies of perceptual grouping report effects of global configuration, consistent with this view of redundancy reduction (Joo et al., 2012). Given the visual system’s adaptive nature, sensitivity to the statistics of natural input may provide a bridge from fundamental mechanisms of suppression and contextual interactions to behaviorally relevant representations of salience in the environment (Kayser et al., 2004; Coen-Cagli et al., 2012), especially as these mechanisms function across different spatial scales (Nurminen and Angelucci, 2014). By representing the influences of visual salience and top-down attention at the earliest stages of cortical visual processing, the visual system is able to achieve a balance between automatic prioritization of local regions throughout the visual field and voluntary guidance based on current goals and tasks.

Grants: This work was supported by the National Science Foundation (Grant BSC-1228526 to FT); the National Science Foundation Graduate Research Fellowship to SP; and the National Institutes of Health (Grant P30-EY008126) center grant to the Vanderbilt Vision Research Center.

Disclosures: None
References


Angelucci A, Bressloff P. Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. *Progress in Brain Research*.

Angelucci A, Bressloff PC. Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. *Progress in Brain Research*.


Beck D, Kastner S. Stimulus context modulates competition in human extrastriate cortex. *Nat Neurosci*.


Kastner S, O’Connor DH, Fukui MM, Fehd HM, Herwig U, Pinsk MA. Functional
SALIENCE AND ATTENTION IN THE EARLY VISUAL SYSTEM


Nelson JI, Frost BJ. Orientation-selective inhibition from beyond the classic visual...


Schallmo M-P, Grant AN, Burton PC, Olman CA. The effects of orientation and attention during surround suppression of small image features: A 7 Tesla fMRI study. 16: 19–21, 2016.


Figure 1. Annotated sample displays for both experiments. Feature-contrast salience is defined by orientation in A and by drifting motion direction in B (indicated by arrows, which were not present during the experiment). Dotted circles (likewise not present during the experiments) indicate the two locations at which the salient patch could appear and to which spatial attention could be directed. While the task-relevant contrast decrement occurred at both of these locations, the observer was instructed to only perform the detection task on one patch, as indicated throughout each block by a small cue. In these examples, if the participant was cued to attend to the black cue (as labeled), in A the left patch is attended/salient, while the right patch is unattended/non-salient; in B, the left patch is attended/salient, and the right patch is unattended/salient. Gabor patch edges were Gaussian blurred and also spatially separated by a gap of 0.8°, yielding an effective spatial separation of approximately 1.8°.

Figure 2. Results of Experiment 1 for the four salience/attention conditions. A) Magnitude of the attention and salience effects across ROIs, computed as difference in percent signal change. Dots show the effect in individual subjects, and error bars depict one standard error (across subjects). In the LGN, only attention significantly modulated BOLD responses; neither the main effect of salience nor the interaction effect were significant. A significant effect of salience first emerges in V1 and is evident in each cortical ROI. Attention also modulated BOLD responses in V1 through hV4, but the two effects do not reliably interact in any region. B) The top panel shows mean ROI time-courses time-locked to the beginning of each experimental block, which lasted 8 TRs and is demarcated by dotted lines. The bottom panel shows the same data averaged across the block, offset to account for hemodynamic lag, and normalized by subtracting the mean response of the 2 TRs immediately preceding the block. Error bars on block averages depict one standard error in the mean BOLD response across subjects.
Figure 3. Examples of average fMRI time courses from representative subjects in Experiment 1 (A, B) and Experiment 2 (C, D). Error bars depict +/- 1 SEM for each experimental condition, and each of the four panels shows data from a different individual. As expected, there is some variability between subjects; fMRI responses in the LGN are also more variable than those in early visual areas, partly because of the LGN's smaller size and the presence of greater physiological noise in midbrain structures.

Figure 4. Magnitude of attention and salience effects plotted as a function of ROI size. We calculated the difference in BOLD response for attended minus unattended conditions and salient minus non-salient conditions, for a wide range of ROI sizes, ranging from 2-60 maximum voxels per hemisphere in the LGN and from 10-350 voxels per hemisphere in individual cortical visual areas. Error bars indicate one standard error across subjects. Our findings of positive salience and attention effects in each cortical ROI are highly consistent across a wide range of ROI sizes.

Figure 5. Mean BOLD amplitudes in Experiment 2, in which salience was defined by the direction of drifting motion of Gabor patches. A) Attention and salience effects (percent signal change difference) across ROIs, with dots plotting individual subjects’ results and error bars showing one standard error across subjects. B) Results in each condition are averaged across the stimulus block and normalized by a pre-stimulus window for each condition. The pattern of results follows that of the first experiment: attentional enhancement is evident in each ROI, including the LGN, while salience modulated only cortical mean BOLD responses. In every region studied, the effects of attention and salience did not significantly interact.
Figure 6. Individual subject ANOVA results for Experiments 1 (left) and 2 (right). Here, each experimental session was analyzed independently, with experimental run as the repeated measure. Filled circles indicate significant effects of attention (black), salience (grey), or their interaction (cross) at p < 0.05 level. Subject labels (S1-S6) are arbitrary and unmatched between the two experiments. While many subjects exhibit significant main effects of attention and salience, the interaction of these two factors is not significant in any ROI in any subject.
Experiment 1: Orientation

Experiment 2: Motion Direction
Experiment 1: Orientation

![Graph A](image)

Experiment 2: Motion

![Graph B](image)
Attention modulation

Salience modulation

% Signal Change

# Voxels selected per hemisphere
A

Attention Effect

Salience Effect

B

attended salient  
attended nonsalient  
unattended salient  
unattended nonsalient
<table>
<thead>
<tr>
<th>LGN</th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
<th>V4</th>
<th>LGN</th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
<th>V4</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>●●●●</td>
<td>●●●●</td>
<td>●●●●</td>
<td>●●●●</td>
<td>S1</td>
<td>●●●●</td>
<td>●●●●</td>
<td>●●●●</td>
<td>●●●●</td>
</tr>
<tr>
<td>S2</td>
<td>●●●●</td>
<td>●●●●</td>
<td>●●●●</td>
<td>●●●●</td>
<td>S2</td>
<td>●●●●</td>
<td>●●●●</td>
<td>●●●●</td>
<td>●●●●</td>
</tr>
<tr>
<td>S3</td>
<td>●●●●</td>
<td>●●●●</td>
<td>●●●●</td>
<td>●●●●</td>
<td>S3</td>
<td>●●●●</td>
<td>●●●●</td>
<td>●●●●</td>
<td>●●●●</td>
</tr>
<tr>
<td>S4</td>
<td>●●●●</td>
<td>●●●●</td>
<td>●●●●</td>
<td>●●●●</td>
<td>S4</td>
<td>●●●●</td>
<td>●●●●</td>
<td>●●●●</td>
<td>●●●●</td>
</tr>
<tr>
<td>S5</td>
<td>●●●●</td>
<td>●●●●</td>
<td>●●●●</td>
<td>●●●●</td>
<td>S5</td>
<td>●●●●</td>
<td>●●●●</td>
<td>●●●●</td>
<td>●●●●</td>
</tr>
<tr>
<td>S6</td>
<td>●●●●</td>
<td>●●●●</td>
<td>●●●●</td>
<td>●●●●</td>
<td>S6</td>
<td>●●●●</td>
<td>●●●●</td>
<td>●●●●</td>
<td>●●●●</td>
</tr>
</tbody>
</table>

- ● attention
- ●● salience
- ●●● interaction