**Neural Substrates of Attention and Awareness**

**First Year Project**

Daniel Birman

Adviser – Justin Gardner

Second Reader – Tony Norcia

## Introduction

In everyday life we feel a direct and undeniable connection between attending to something and our clear awareness of it. Despite this there exist a variety of laboratory situations in which attention and awareness appear to diverge. This apparent disconnect between experience and experimental findings has fueled a debate about whether *selective attention* and *awareness* are dissociable (Koch & Tsuchiya, 2007). In scene recognition experiments participants are able to identify the content of peripheral stimuli despite attending to a demanding fixation task (Li, VanRullen, Koch, & Perona, 2002). In contrast, many studies have found that participants are unable to respond about unattended stimuli, in particular for simple features such as shapes and colors (Mack, Arien & Rock, Irvin, 1998). One possible interpretation for the variability in results in these tasks is that stimulus features may be interacting in unexpected ways. In this view, only certain types of attentional modulations would modify the neural activity responsible for scene recognition. In line with this hypothesis is the finding that attending to motion does result in diminished awareness of scene gist (Cohen, Alvarez, & Nakayama, 2011). We propose that this feature-specific hypothesis is a property of the hierarchical organization of visual cortex. Our hypothesis is that modifying a neural representation, through spatial or feature-based attention, ultimately feeds forward to downstream visual areas. The impact of these changes will depend on whether they cause constructive or destructive interference relative to a particular task. We propose to test this hypothesis by investigating whether contrast and motion coherence interfere in an asymmetrical manner, and whether this asymmetry is well predicted by neural responses in representative areas. Our prediction is that attention to contrast will be detrimental to the perception of motion, but not vice versa, due to the largely feed-forward organization of visual cortex.

A number of theories have now been put forward about the functional implementation of top-down attention. In the early visual areas in humans for example there is evidence that spatial attention acts as a form of response gain (Itthipuripat, Ester, Deering, & Serences, 2014) or a kind of selective gating of activation (Pestilli, Carrasco, Heeger, & Gardner, 2011). But little is known about how attention impacts visual perception of unattended stimuli. We know that changes in population responses in visual cortex correlate with perception (Cutrone, Heeger, & Carrasco, 2014), so we might expect that the influence of spatial attention on unattended regions should also predict changes in visual perception in those areas. Note that this view makes no explicit claim about how unattended features or spatial locations will be modified, only that any change in their activity will manifest as a change in visual perception. If attention-related modulations introduce destructive noise, then visual perception will be weakened. On the other hand, if the modulation is constructive, visual perception could be improved. Given our knowledge of the hierarchical organization of visual cortex (Grill-Spector & Malach, 2004) it is plausible that modulatory effects will manifest in asymmetrical ways. For example, because projections from V1 to hMT+ are stronger than feedback connections we might expect that only modulations of V1 will impact MT, but not vice versa. This is analogous to saying that attention to features represented by V1 will impact motion perception, but attention to motion will not impact features represented by V1 (under the assumption that MT represents motion). We tested this general hypothesis by asking participants to respond about unattended features while we recorded neural responses. Our prediction

was that the neural representation of motion coherence is affected by attention to contrast, and that the effect is a destructive interference. To summarize: in our task we expect that attending to contrast will reduce performance on motion coherence discrimination, but that attending to motion coherence will have no effect on contrast discrimination.

Our behavioral task design resembles work that has been done on "dual task" experiments, in particular the extensive line of literature on inattentional blindness (Mack, Arien & Rock, Irvin, 1998). This is the same line of experimental results that have found differences in scene gist recognition under varied experimental conditions (Cohen et al., 2011; Li et al., 2002; Mack & Clarke, 2012). Measuring behavioral responses for "unattended" stimuli is difficult. Many dual task designs use stimulus novelty to assess recognition or discrimination performance for unattended stimuli. Because of this, they are usually limited to collecting a single critical trial. Our goal here is to assess whether neural activity in precise anatomical regions is sufficient to explain the range of behavioral performance. This within-subject design requires precise estimates of subject-level performance, which are impossible to obtain in a single trial. With this in mind we chose to use a classic signal 2-alternative forced choice (2-AFC) discrimination task, using miscued trials to assess performance on the un-cued feature. This approach may introduce its own issues if participants 'split' their attention between stimulus features. To mitigate this confound we control for task difficulty in our design. We also measured a control condition in which cueing was 100% predictive of response category, so that differences due to attention 'splitting' could be easily assessed.

We view attention in visual space as a local context-dependent modulation of neural activity in a precise region of visual cortex. Visual perception appears to be a downstream property of neural activation that has a direct dependence on activations in visual cortex. These assumptions predict that asymmetries in how attention interacts with awareness should be well predicted by the topology and interconnections of visual cortex. For example, there is a known asymmetry between scene gist recognition being abolished by motion (Cohen et al., 2011) but not by other demanding attentional tasks (Li et al., 2002). Our hypothesis predicts that motion representative neural activity should feed forward into the areas that represent scene gist, but not vice versa. We tested this general hypothesis with a more basic task involving judgments of contrast and motion coherence. Our prediction for our task is that attention to contrast will affect motion coherence, but not vice versa. In addition, we expect that this relationship will be precisely explained by the variability, due to attentional cueing, in neural responses in contrast and motion responsive regions of visual cortex. Understanding this process and characterizing this asymmetrical relationship will help us understand the role of attention in visual cortex and its interactions with downstream processes such as visual awareness. In addition, characterizing a general architectural constraint, such as the one we have outlined here, predicts similar hierarchical influences on memory, cognitive control, learning, and on any other downstream processes.

## Methods

### Subjects

Four human subjects (all male, ages 24-34) participated in the experiment. All subjects performed the behavioral experiment and one participant performed the functional MRI experiment. All subjects performed one training session (30 – 120 min) to become accustomed to the task, four to eight control runs (65 trials each), and between 5 and 18 task runs (100 trials each, 15% miscued). One subject performed a retinotopic mapping (60 min, ten 4 minute scans) and three sessions of the main experiment (120 min each, consisting of ten, eight, and eight 7 min scans respectively, the final scan session included a motion localizer scan of 4 minutes).

**Experimental Task**

Subjects performed a two-alternative forced choice (2-AFC) feature discrimination task. On each trial participants were shown two patches of dots and asked to report which had a higher contrast or motion coherence. Participants were cued on each run (65-100 trials during behavior or ~47 trials during scanning) to attend selectively to one of the two features. Each dot patch was shown for 750 ms and followed by a 250 ms mask, generated by flashing a random checkerboard pattern at 55% contrast. The checkerboard mask changed pattern pseudo-randomly at 40 Hz. A random inter stimulus interval (ISI) followed the mask for between 200 and 500 ms after which the fixation cross turned white, indicating that the subject should make their response. Participants had 1 second to respond. Each trial was followed by a random inter-trial-interval of 300 to 500 ms. To improve estimation of the hemodynamic response the trial timing was modified during scanning. Stimulus length: 750 ms, mask: 250 ms, ISI: 200-1000 ms, resp: 1000 ms, ITI: 2000-10000 ms. Throughout the experiment participants fixated a black central cross (1 deg x 1 deg visual angle, 1 pixel wide).

The dot patches appeared left and right of fixation, extending from 3.5 to 11 degrees horizontally and from -5 to 5 degrees vertically. The patches were displayed on a gray background (50% luminance) on a monitor with a linearized luminance scale. Each patch contained 1000 dots, half of which were darker than the background and half of which were brighter. Each dot was 4 x 4 pixels on the screen. The luminance difference between the dots and the background was defined as $\frac{C}{2}$, where $C$ is the contrast (0 to 1) on the current trial. A percentage of the dots, $M$, moved horizontally either right or left (randomly chosen on each trial) while the remaining $1 - M$ dots had random angles. $M$ therefore reflects the motion coherence of the dot patch on a scale of 0 to 1. Dots moved at 3.25 degrees / s.

We separated the experimental runs into control and task runs. Participants were cued to the current feature at the start of the run. During control runs participants were cued to attend to either motion or contrast and they responded about the same cued feature. On task runs participants were occasionally informed, after stimulus presentation, that they should respond about the miscued feature. This post-stimulus cueing was performed by replacing the fixation cross with a letter, ‘C’ indicated that they should respond about contrast, and ‘M’ indicated that they should respond about motion. These task miscued trials occurred 15% of the time and were interspersed pseudo-randomly. On miscued trials the ISI was fixed at 500 ms and the response time was increased to 3000 ms.

Contrast and motion discrimination performance were tested at single pedestal intensities. Contrast was tested at 60% and motion coherence was tested at 10%. The features were crossed such that neither feature was informative about the intensity of the other feature on any given trial. There

was always a difference between both the contrast and motion coherence of the two dot patches, but on a percentage of trials this difference was perceptually invisible. For each feature a 1-up-3-down PEST staircase (Taylor & Creelman, 1967) was used to set the increments in contrast or motion coherence that were added to the pedestal contrast on the target side. The independent staircases balanced task difficulty across the features so that subjects were always performing the task at or near "threshold", eliminating a potential confound with task difficulty between conditions. Threshold was defined as a performance rate of ??. During control runs pedestal values of 20/40/60/80% contrast and 0/10/20/40% coherence were used to allow estimation of the BOLD response across a larger range of feature intensities. On scanning runs each pedestal had an independent staircase and pedestals were pseudo-randomly interleaved across trials.

**Stimulus Presentation**

Outside the scanner the visual stimuli were presented on a ViewPixx 22.5" LCD (VPixx Technologies) with a resolution of 1920 x 1200 pixels and a 100 Hz refresh rate at a distance of 55 cm from the subject's eyes. Inside the scanner subjects used an adjustable mirror system to view an image that was rear-projected onto a fiberglass screen using an Eiki LC-WUL100L projector operating at 1920x1200, 5000 lumens, projected through a neutral density filter at 60 Hz. The projector and LCD screen were calibrated to have linearized gamma scales using a PR650 Spectroradiometer (Photo Research Inc., Chatsworth, CA.). On each trial we dynamically adjusted the 10-bit gamma table to achieve the best luminance resolution possible (maintaining the linearized output) for displaying each dot patch. All stimuli were produced using MATLAB (The Mathworks Inc., Natick, MA, USA) and MGL (http://gru.stanford.edu/doku.php/mgl/overview) using custom scripts accessible online (cohcon.m and mtloc.m, https://github.com/justingardner/grustim).

**Eye Position Measurements**

An Eyelink 1000 eye tracking system (SR Research Ltd., Mississauga, ON, Canada) was used outside the scanner to confirm that subjects maintained fixation throughout the task. Eye tracking was not performed inside the scanner. The Eyelink system recorded corneal reflections of an external infrared light source and tracked the center of the pupil. A brief calibration was performed before each 5-minute run. Eye tracking setup was successful for most sessions, sessions where eye calibration failed were run with the calibration form the previous session. The calibration data was used to perform an affine transformation of the acquired eye tracking data to the position of the eye in degrees of visual angle.
(As of 6/1/15 I haven't had time to add the eye tracking analysis to the results.)
To assess the stability of subject fixation during the experiment we plan to perform a regression analysis predicting eye position using time from trial start, cueing type (contrast or coherence), response type (contrast or coherence), and correct side (left or right). We expect to observe no significant deviations from fixation at any consistent time relative to the trial start or condition. In particular we expect not to observe consistent deviations towards the target side.

**Contrast and Motion Discrimination Functions**

Feature discrimination task performance was evaluated by computing a feature-discrimination function. A feature-discrimination function defined the relationship between the pedestal intensity (i) and the increment in intensity (deltai) required to obtain threshold-level performance. Feature-discrimination functions were computed separately for contrast and motion coherence. For each condition a maximum-likelihood procedure (Wichmann & Hill, 2001) was used to fit subject responses to a Weibull function (Weibull, 1951):

$$p(\Delta i) = \left(\frac{1}{2} - \delta\right)\left(1 - e^{-\left(\frac{i}{\tau}\right)^{k}}\right) + \frac{1}{2} \quad (1)$$

Where $p(\Delta i)$ is the probability of being correct given an intensity increment of $\Delta i$, $\delta$ is the lapse rate, $k$ is the slope of the psychometric function on a log-log axis, $\tau$ is the $\Delta i$ for which the probability correct reaches 63% of the difference between chance and maximal performance, and m is the slope of the psychometric function. Subjects performed on average 4 psychometric function staircases with 50-100 trials each. A minimum of 50 trials were allowed per function, sufficient to estimate the discrimination threshold accurately (Kontsevich & Tyler, 1999). By running multiple staircases we ensured that we were able to capture any drift in threshold performance over time (for example due to motivation). Running multiple staircases also allowed us to directly compute the variability of parameter estimates across runs. Variability can also be estimated by pooling staircases across runs and performing a bootstrap procedure (Wichmann & Hill, 2001).

For each participant we estimated their discrimination functions for control runs, task cued trials, and task miscued trials. Each of these discrimination functions therefore plots the performance $p(\Delta i)$ at all possible threshold values. To make quantitative comparisons between these performance curves they were normalized to the threshold estimated from the control runs, when participants were cued with 100% accuracy about the response feature.

**MRI Acquisition and Preprocessing**

MRI data were acquired on a GE Discovery MR 750 on a Nova Medical 32ch head coil. Retinotopy experiments were collected on a Nova Medical 16ch visual array. For each subject we acquired a high-resolution 3D anatomical image ("canonical anatomy") which was segmented via FREESURFER (http://surfer.nmr.mgh.harvard.edu) to generate white matter and gray matter segmentation (Dale, Fischl, & Sereno, 1999). We collected a single T1-weighted image (MPRAGE TR 7.24 ms, TE 2.78 ms, FA 12°, voxel size 0.9 x 0.9 x 0.9 mm, matrix 256 x 256. Regions of interest were drawn on flattened representations of the cortical surface including the visual areas and the motion sensitive regions that defined hMT+. These regions of interest were constrained to voxels that intersected the gray matter. Analyses were conducted on original untransformed data while flattened representations were used for visualization.

Each functional experimental session consisted of a lower resolution T1-weighted image ("session anatomy") and multiple T2*-weighted functional scans (multiband 8, TR 500 ms, TE 30 ms, flip angle 47°, voxel size 2.5 x 2.5 x 2.5 mm, matrix 88 x 88). Additional information about multiband

(Feinberg & Setsompop, 2013) sequences can be obtained through the Stanford Center for Cognitive and Neurobiological Imaging (http://cni.stanford.edu/wiki/MUX_EPI). An automated procedure was used to find the best affine transform to align the session anatomy to the canonical anatomy (Nestares & Heeger, 2000). The functional scans were aligned to the session anatomy directly using the qform coordinates obtained from the scanner. Retinotopic mapping was performed using a T2*-weighted functional scan (multiband 2, TR 1400 ms, TE 30 ms, flip angle 55°, voxel size 2.5 x 2.5 x 2.5 mm, matrix size). Oblique slices were chosen to maximally cover the occipital visual areas, approximately perpendicular to the calcarine sulcus. For all subjects our functional sequences achieved full brain coverage.

fMRI analysis was performed with a custom pipeline using MATLAB, mrTools (http://gru.stanford.edu/doku.php/mrTools/overview), and custom scripts accessible online (https://github.com/dbirman/att_awe).

**Retinotopy**

Visual field maps were drawn based on retinotopy sequences collected in a separate scanning session. High-contrast radial checkerboard patterns were presented either as an expanding or contracting ring or a 90° rotating wedge. Each scan consisted of 10.5 cycles (24 s per cycle) of the ring expanding/contracting or the wedge completing a full rotation with a sampling rate of 17 volumes per cycle (178 volumes per scan). In addition four presentations of a sweeping bar stimulus were made. Each session therefore consisted of two scans of the ring stimulus (one expanding, one contracting), four scans of the wedge stimulus (two each clockwise and counter-clockwise), and four scans of the bar stimulus. A generative model of voxel responses (the Population Receptive Field model, Dumoulin & Wandell, 2008) was fit to each voxel, identifying the Gaussian parameters (x, y, sigma) that best fit the recorded response data. Visual fields were then defined according to established criteria (Wandell, Dumoulin, & Brewer, 2007).

**Functional Motion Localizer**

In addition to retinotopic mapping of the visual field we identified regions that were responsive to optic flow motion with a functional localizer sequence (Huk, Dougherty, & Heeger, 2002). This localizer sequences oscillates every 12 s between an optic flow stimulus (with direction reversals every 0.5 s) and a noisy motion stimulus, where each optic flow dot's motion vector in 3-d space is rotated randomly on each frame. We identified hMT+ in our retinotopic maps via established procedures (Wandell et al., 2007), as well as with our functional localizer. We identified voxels that showed a correlation greater than 0.1 with a sinusoidal model response and excluded voxels in the early visual areas (V1-V4). This procedure identified a patch of voxels that selectively responded to coherent motion. We restricted subsequent analysis to the conjunction of voxels between retinotopically and functionally defined hMT+.

**Feature-response Functions**

To compute the feature-response functions, a deconvolution analysis was used to determine the mean hemodynamic response to each dot patch in the contralateral visual cortex, for details see: (Gardner et al., 2005). The average time-course in each visual area for each grating location was

computed and the response following stimulus presentations for 15 s was calculated, assuming linear summation for responses that temporally overlapped. These responses were calculated separately for each combination of feature (contrast, motion coherence) and cueing condition (cued, miscued) at every intensity increment, rounded to the nearest 10%. This resulted in 36 total conditions (contrast: 8 intensities x 2 cueing + coherence: 10 intensities x 2 cueing). A gamma function was fit to each de-convolved response and the amplitude of this function determined the magnitude of response. These response magnitudes were then plotted as a function of stimulus intensity to yield the feature-response function for each visual area and cue condition. These feature-response functions were then parameterized to the Naka-Rushton equation.

$$R = \frac{R_{\max} F^n}{(F^n + F_{50}^n)} + b \quad (2)$$

Where R represents the BOLD response, $F$ is the feature intensity, $b$ is the background response, and $n$ is the exponent of the power function. $R_{max}$ and $F_{50}$ are free parameters that were fit for each feature (contrast, coherence) and condition (cued, miscued).

**Modeling Discrimination as Neural Responses**

Discrimination performance is a direct function of neural responses. One of our goals is to understand whether the responses in V1 and hMT+ are sufficient to explain the discrimination performance that we observe. To test this we will model our discrimination task as a signal discrimination task, using $d'$ as a measure of performance. It follows then that the discrimination threshold, $d'$, should also correspond to the following equation:

$$d' = \frac{R(i + \Delta i) - R(i)}{\sigma_R} \quad (3)$$

Where R is the response to a stimulus intensity $i$, estimated from equation 2. $\sigma_R$ is the variance on the response function. An alternative model to this is that attention acts as a modulator of neural signals as part of a downstream selection process (Pestilli et al., 2011), we will also test whether this model is a better explanation of our cueing results.

**Results**

Participants in our experiment performed a 2-AFC choice discrimination. Subjects compared two patches of dots and responded about which patch they perceived as having the higher contrast or motion coherence, depending on the trial type. Each participant was cued at the start of the run about which feature they should attend to. Participants performed control runs during which they were only asked to discriminate one of the two features, as well as task runs where on 15% of trials the subjects were told, after stimulus presentation, to respond about the un-cued feature. Subjects performed training runs for one to two hours or until their contrast and coherence thresholds stabilized. We found that subjects were able to immediately perform the contrast task at near maximal performance, but took between 30 and 120 minutes to learn and perform the coherence discrimination task successfully.

Subjects then completed as many runs of as possible, alternating between control runs where they were told that no miscued trials would occur, and task runs that included miscued trials. In total subjects performed 793, 821, 679, and 681 trials (subjects 300, 302, 25, 21, respectively). Subject 300 additionally performed ~1300 trials during scanning in 26 runs, plus a four minute motion localizer.

Figure 1 shows the estimated thresholds for each participant during each of the different run conditions estimated across all of the staircases that were run for that participant. We expected two notable patterns to emerge in these data: (1) performance on each task should be equal during control runs and task cued trials, when participants are attending to one feature and responding about that feature, (2) performance on task miscued trials should be significantly lower (i.e. a higher discrimination threshold) than in either of the other conditions. We expected (1) to be true only if participants were not "splitting" their attention between tasks during the task cued condition. The weaker performance of subjects 21 and 302 on task cued trials for coherence, for example, suggests that they did split their attention to some extent. We expected (2) to be true based on the large wealth of dual-task experiments that show that attention enhances performance on visual perception tasks. Our results corroborate these findings and show that when participants are miscued about the relevant response type this reduces their ability to respond correctly. This result is most pronounced for coherence, but quantitative comparisons cannot be made between tasks in this figure. We chose to normalize the results by the performance in the control runs to better estimate the relative performance on the task runs.

To compare the normalized results we plot in Figure 2 the feature-discrimination functions for subject 300. These functions are a measurement of performance (% correct) at increasing differences of feature intensity between the two dot patches. The "threshold" in Figure 1 corresponds to a single point along this function. We found that there was a variable but consistent effect showing that, across subjects (Figure 2, inset), the normalized performance for contrast when miscued was within the same range as the performance when cued or during control runs. In contrast, performance on motion coherence discrimination suffered dramatically during the task miscued trials. We also observed a difference in performance between the task cued and control performance for motion coherence. We focused the remaining analysis on the performance curves of subject 300 who also performed the functional imaging experiment. Subject 300 shows the typical feature-discrimination function shape observed across the experimental population. Performance on miscued contrast trials does not suffer as much as performance on miscued coherence trials. For subject 300 we did not observe any change in cued coherence trials between the control and task runs.

The eye tracking analysis hasn't been completed (as of 6/1/15).

Based on our knowledge of the functional architecture of early visual cortex (Grill-Spector & Malach, 2004) we expected to find that response amplitudes would be modulated in V1 for contrast intensity, but not for motion coherence. We expected the opposite relationship in human MT, or that an interaction would occur where both contrast and motion coherence would contribute to response amplitudes. In addition, our prediction was that the modulation of contrast response in V1 would be sufficient to account for the behavioral effect of contrast cueing. We expected that any downstream response modulations would be additionally uninformative, simply reflecting the feed-forward activity

from V1. Likewise, we predicted that changes in the response magnitudes in MT would be sufficient to account for the behavioral effect of motion coherence cueing. We also expected to see an asymmetrical relationship—where cueing to contrast would have an effect on response amplitudes in MT but no reciprocal effect of coherence on response amplitudes in V1. This prediction is based solely on the architecture of the early visual cortex, where the vast majority of neural connections are feed forward.

To assess the responses of V1 and MT to stimulus of varying contrast and motion coherence intensities we first isolated the visual regions. In Figure 3 we present the results of our functional and retinotopic mappings. We found perfect overlap between the functionally defined hMT+ and the retinotopically defined map and restricted voxels to the overlap for subsequent analysis. We performed a deconvolution analysis (see Methods) and present the resulting $R^2$ values in Figure 4. The rectangular stimulus (here presented to the right visual field) appears in a retinotopically correct subsection of V1. We found voxels that were sensitive to our stimulus in all of the visual field maps, but we focus our subsequent analysis only on V1 and hMT+.

Figure 5 shows an early view of the neural data regarding contrast and motion coherence representations in visual cortex. So far we have only analyzed the left visual cortex using the stimulus presentations to the right visual field. These early results do indicate sensitivity to motion coherence and contrast in V1 and MT, as well as possible sensitivity to the cueing conditions. We plan to extend these analyses significantly in the future to obtain a clear view of how visual cortex is representing these features. We plan on analyzing both visual cortices separately, then combining data if the HRF fit results are similar across the hemispheres. We also plan to investigate which model (Figure 6) best fits the cueing data to understand what form attentional modulation takes in these visual areas. This analysis can also be repeated in the other visual fields, but we will restrict our initial analysis to V1 and hMT+. In addition, we will look at whether there is sufficient neural data to estimate the hemodynamic response during catch trials, where participants may have been able to attention retroactively to their memory of the stimulus.

The modeling analysis hasn't been completed (6/1/15). The possible alternative models for how attention might modulate visual cortex are presented in Figure 6. We plan to use equation 3 to map the neural responses onto the behavioral responses to estimate whether the magnitude of change in the neural response is directly related to improved behavioral performance, or whether there is a secondary mechanism involved (see e.g. Pestilli et al., 2011). Our expectation is that spatial feature attention in V1 for contrast is similar to spatial attention and that our results will be well fit by a model of efficient selection on the V1 activity. We don't have any expectation yet about how the results will map in hMT+ for motion coherence, but we do expect that both contrast and motion coherence will impact responses in that region, and therefore have an impact on performance. This assumption is based on our behavioral data where we observed an interaction where cueing to contrast has a strong effect on the ability to perform motion coherence discriminations.

**Discussion**

Visual cortex in humans is organized in a hierarchical manner, from basic feature representations up to more complex representations of objects, motions, and scenes. The majority of activity in visual cortex proceeds in a feed-forward manner, from input via the lateral geniculate nucleus to V1 and upwards (Grill-Spector & Malach, 2004). Visual cortex also receives local inputs, inputs from other parts of visual cortex, and top-down modulation from other cortical regions. This feed-forward connectivity predicts that there will be asymmetries in how a local modulation of activity will spread through cortex. Modulations of V1, for example, are likely to be pushed forward into other visual regions. In contrast, modulation of downstream regions such as hMT+ are unlikely to spread back to the earlier visual regions. At the most extreme we can hypothesize that feature specific modulations of a cortical region, such as those introduced by covert feature-based attention, might introduce noise into other visual regions. We suspect that these mechanics may be responsible for a previously unexplained asymmetry in human performance on dual-tasks involving scene gist recognition. It is known that under some demanding viewing conditions subjects can recognize scene gist for an unattended scene at the same accuracy as under free-viewing (Li et al., 2002). This relationship breaks down in certain specific circumstances, such as when participants attend to motion (Cohen et al., 2011). This study was designed to test the hypothesis that hierarchical organization may be sufficient to explain similar performance asymmetries found for judgments of contrast and motion coherence. Explaining this apparent asymmetry would fill a gap in our understanding of how selective attention impacts visual awareness.

We found that performance did not suffer equally when participants were asked to judge contrast while cued to coherence compared to the reverse. Discrimination performance for coherence judgments was markedly reduced, while on average subjects showed only a small change in contrast discrimination when cued to motion coherence. A portion of this effect may be due to attention "splitting" between the main motion coherence task and the un-cued contrast discrimination task. We observed that there was a difference between performance on motion coherence discrimination between these task runs, which include miscued contrast trials, and control runs which included only motion coherence trials. This difference may be due in part to the relative ease of performing the contrast discrimination relative to the motion coherence discrimination. Participants did not require training to perform contrast judgments, but all subjects required a minimum of thirty minutes and up to two hours of training before motion coherence discrimination thresholds stabilized. It appears that despite controlling for task difficulty (in terms of percent correct) we could not control for the relative ease of performing each type of discrimination. By using a modeling approach to link behavior and neural activity directly we believe we can avoid any confounds caused by this difference in task difficulty.

We interpret our behavioral findings as support for the overall hypothesis that hierarchical organization in cortex may be responsible for asymmetries in judgments of unattended stimuli. To fully characterize this relationship we also recorded neural activity from the visual cortex. We isolated activity from two visual regions, V1 and hMT+, that are likely candidates for the neural correlates of visual perception for contrast and motion coherence respectively. Other studies have identified V1 activity as sufficient to fully explain behavioral performance on contrast judgments (Itthipuripat et al., 2014; Pestilli et al., 2011). Because similar studies have not been performed for discrimination of motion coherence

we plan to test our models to establish this fact for hMT+ as well (this part of the project is in progress). We also plan to test whether our hierarchical model.

Our full results, including the complete neural data analysis and the model analysis, remains to be completed, but the initial behavioral findings point towards the hypothesis we outlined. We suggested that taking into account the hierarchical nature of visual cortex might be sufficient account for contradictory results in the inattentional blindness literature. We tested this hypothesis using contrast and motion coherence, two features represented in hierarchically distinct parts of cortex. We found that our behavioral results are consistent with the overall hypothesis. It remains to be seen whether the full model analysis as well as our neural data also point towards a cohesive view. In addition, if this is a general feature of hierarchically organized cortex we might expect a number of unexpected interactions to exist between stimulus dimensions. Future work will need to extend these findings into other feature domains and test whether this is a general principle of cortical connectivity. If our results extend into other feature domains then this is strong evidence that attention and awareness are more complex than expected. The effect of 'attention' to any specific location or feature cannot be predicted without full knowledge of the context. This includes the specifics of the attended stimuli and of all unattended stimuli. This more complex picture of attention and awareness is not necessarily less understandable, but it requires taking into knowledge of the neural dynamics and cortical connectivity. Further testing of this attention-awareness hypothesis in other domains will help establish whether attention always interacts with perception in context specific ways, or whether there are regions of cortex where local connectivity plays a less pronounced role.

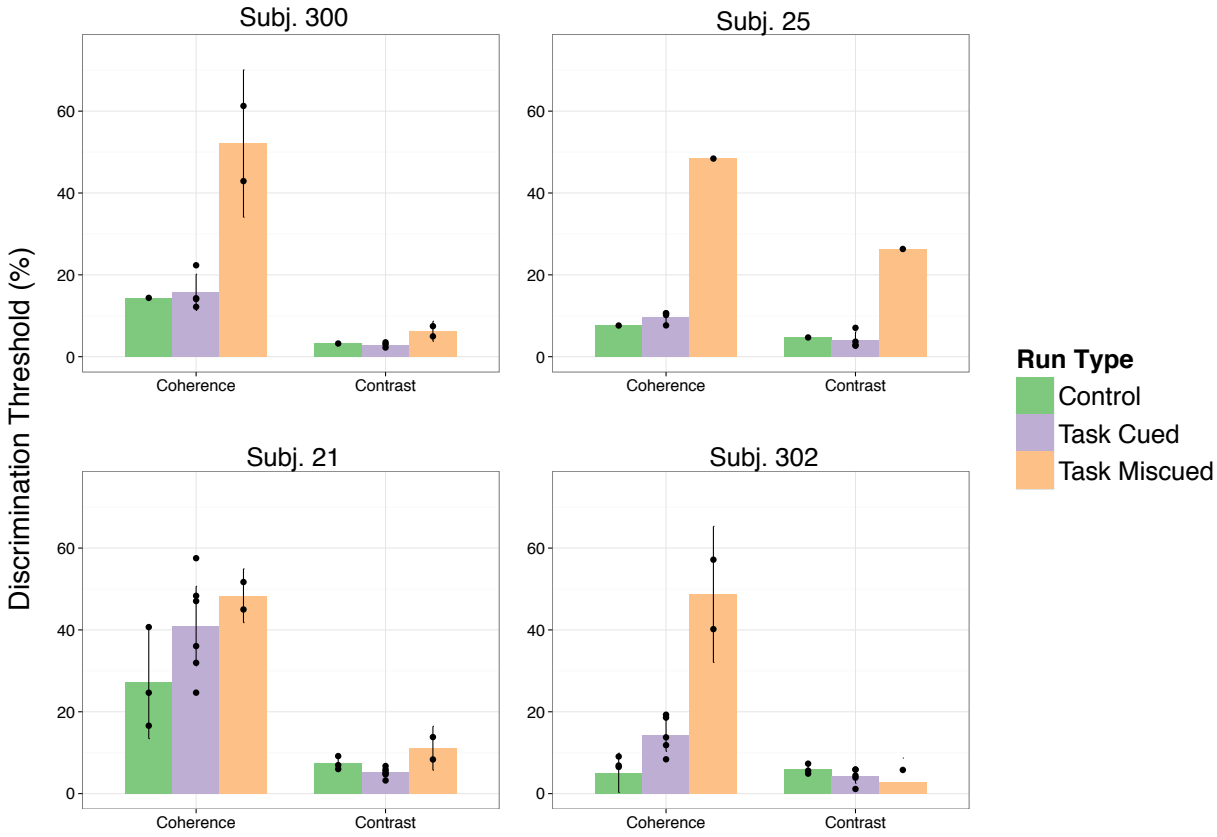


**Figure 1**

Estimated Discrimination Thresholds Across Participants. Each participant was shown several hundred trials in which two moving dot displays varied in both contrast and coherence. On each trial they made a discrimination judgment about which display had a higher feature intensity. On control runs they were cued to attend to the same feature that they discriminated. On task runs they were mostly cued to the correct feature, but 15% of trials were "miscued": participants had to respond about the unattended feature. We fit each participant's raw performance at different feature intensity differences to a Weibull function (equation 1) and extracted "threshold" performance, corresponding to 82% correct. Threshold performance is plotted here by feature and condition for each subject individually. We observed a relatively stable pattern where miscued judgments about motion coherence were much more difficult (i.e. a larger difference in feature intensity was required for the same performance) than miscued judgments about contrast. We expected that if participants were focusing their attention entirely on the cued feature, as directed, their performance on the control and task cued trials would be identical. This was the case for contrast, but not for motion coherence where we found that performance was better during control runs compared to task cued runs. We directly compared discrimination across features by normalizing these data to the control runs, see Figure 2 and text for more details.

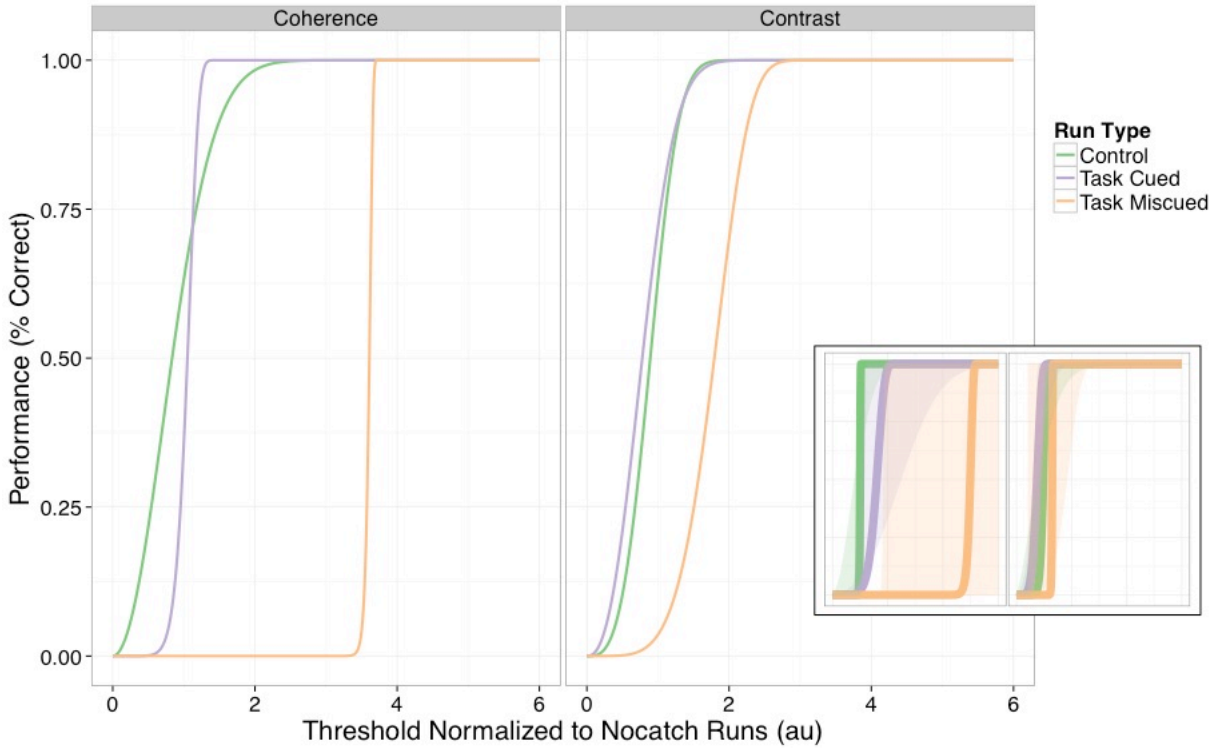


**Figure 2**

Normalized Feature-Discrimination Functions. Feature discrimination function fits (equation 1) are shown for subject 300 and averaged across participants in the inset. Performance was relatively consistent for all participants in the contrast discrimination trials for both control runs, task cued trials, and task miscued trials. In contrast, motion coherence discrimination suffered during task miscued trials. Performance on motion coherence also suffered a smaller amount during task cued trials, suggesting a possible confound of task difficulty (see the text for details). Discrimination function slopes were relatively consistent across both tasks and conditions, so we compared only the estimated thresholds to determine effect size. We computed a mixed-effects analysis predicting normalized thresholds from condition and task and their interaction. Thresholds were normalized within subject to account for individual variability. We found a significant interaction between a quadratic term estimating that task miscued thresholds were significantly higher than both control and task cued thresholds, beta = .82, $t(59) = 2.08$, $p = .042$. The results obtained for subj. 300 during scanning were qualitatively similar to the results obtained during the behavioral runs.

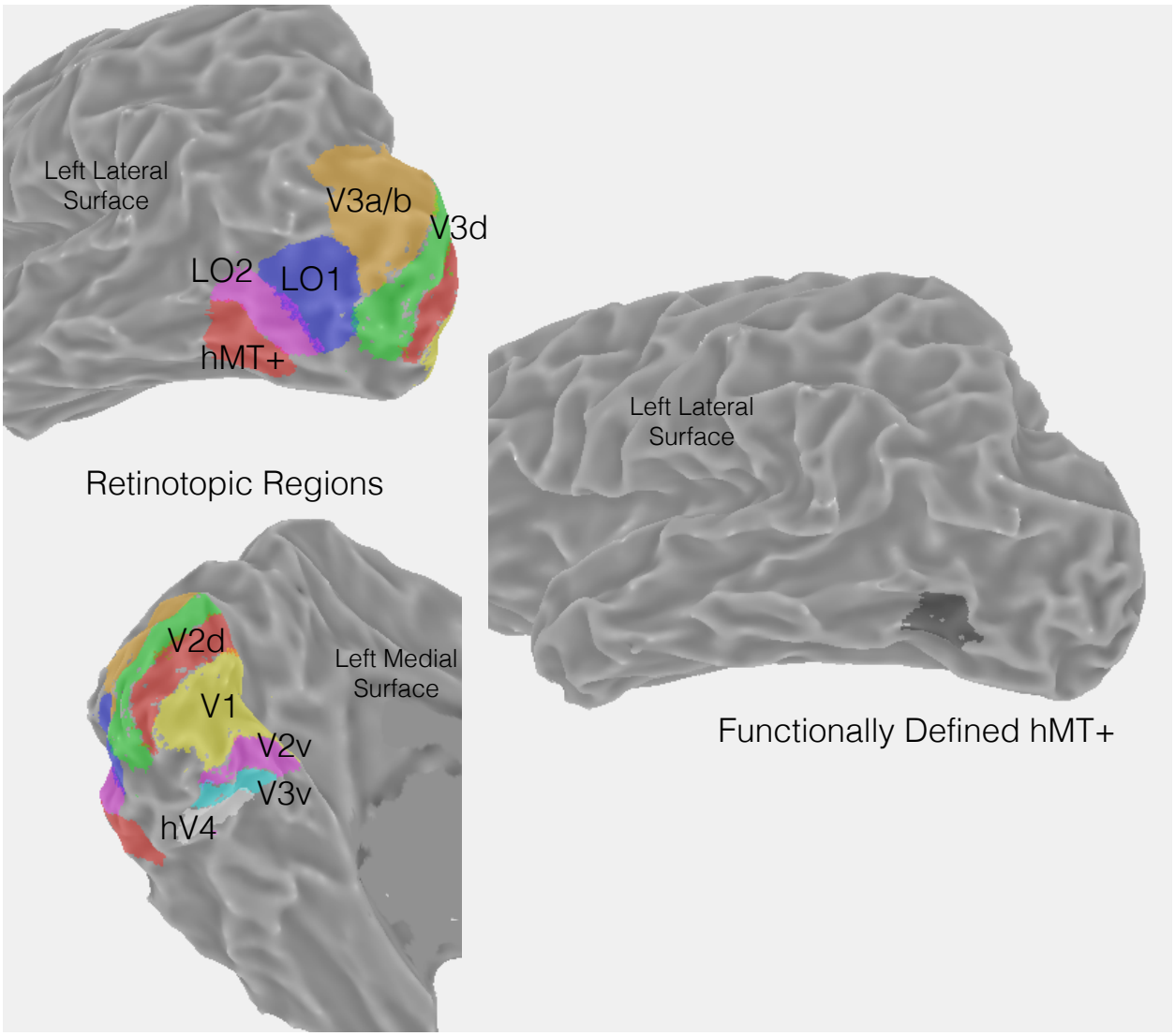


**Figure 3**

Surface Visualization of Left Hemisphere Regions of Interest (ROIs). Regions were first defined retinotopically on a flattened visualization of the left hemisphere occipital cortex according to established criteria (Wandell et al., 2007). First, a population receptive field model was fit to the voxels in occipital cortex. Second, visual areas were defined according to the polar angle flips at visual area boundaries. Voxels were then projected onto an inflated surface for visualization. Area MT was also defined according to a functional localizer by correlating BOLD activation to a 12 s ON, 12 s OFF motion stimulus with a sinusoidal wave. Area hMT+ was identified by restricting the ROI to voxels with a correlation greater than 0.1 and excluding voxels in the retinotopically defined early visual areas (V1, V2, and V3). Functionally defined hMT+ overlaps completely with the retinotopically defined area.

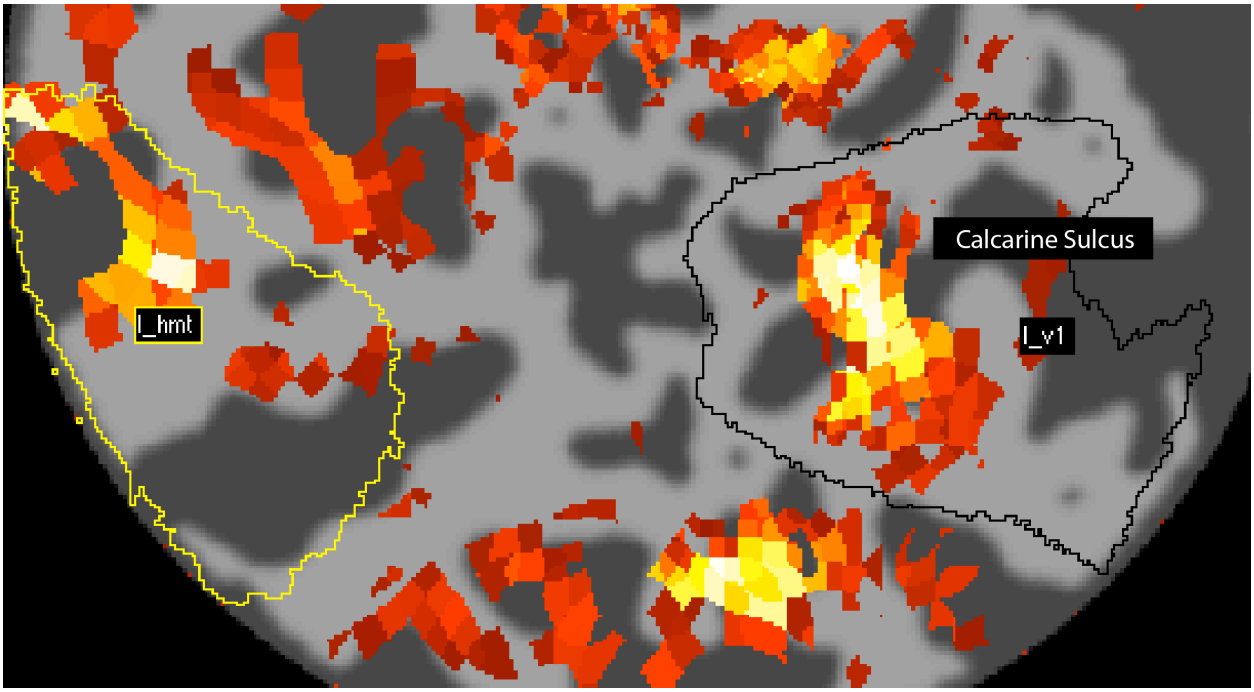


**Figure 4**

Event-Related Deconvolution Variance Explained Displayed on Flattened ROIs. After performing the event-related deconvolution (see Methods) all voxels surviving an $R^2$ threshold of 0.2 are shown relative to the retinotopically defined V1 and hMT+ ROIs in a left hemisphere occipital cortex flat map. Heat map colors indicate $R^2$ values across the region. Note that the screen display was a rectangle approximately 3 to 11 degrees eccentric from fixation, extending 5 degrees above and below the midline. We see active responses as expected in V1 and both dorsal and ventral V2 and V3, as well as a patch of responses in hMT+. We also observed retinotopically consistent activation in V4, LO-1 and LO-2, and V3a/b.

**Figure 5**

[Figure 5 has technical issues, the final version couldn't be generated on time, but I will send an updated version when possible. The version I was going to include didn't include the full dataset and wouldn't be useful for interpretation.]

Feature Response Functions. We computed the BOLD response in the average of responsive voxels in V1 and hMT+ and then de-convolved the response to each stimulus intensity. The de-convolved responses were refit to a difference of gamma function and the amplitudes are plotted above. These amplitudes values were then fit with the Naka-Rushton equation (equation 2) to reduce noise in the estimate.

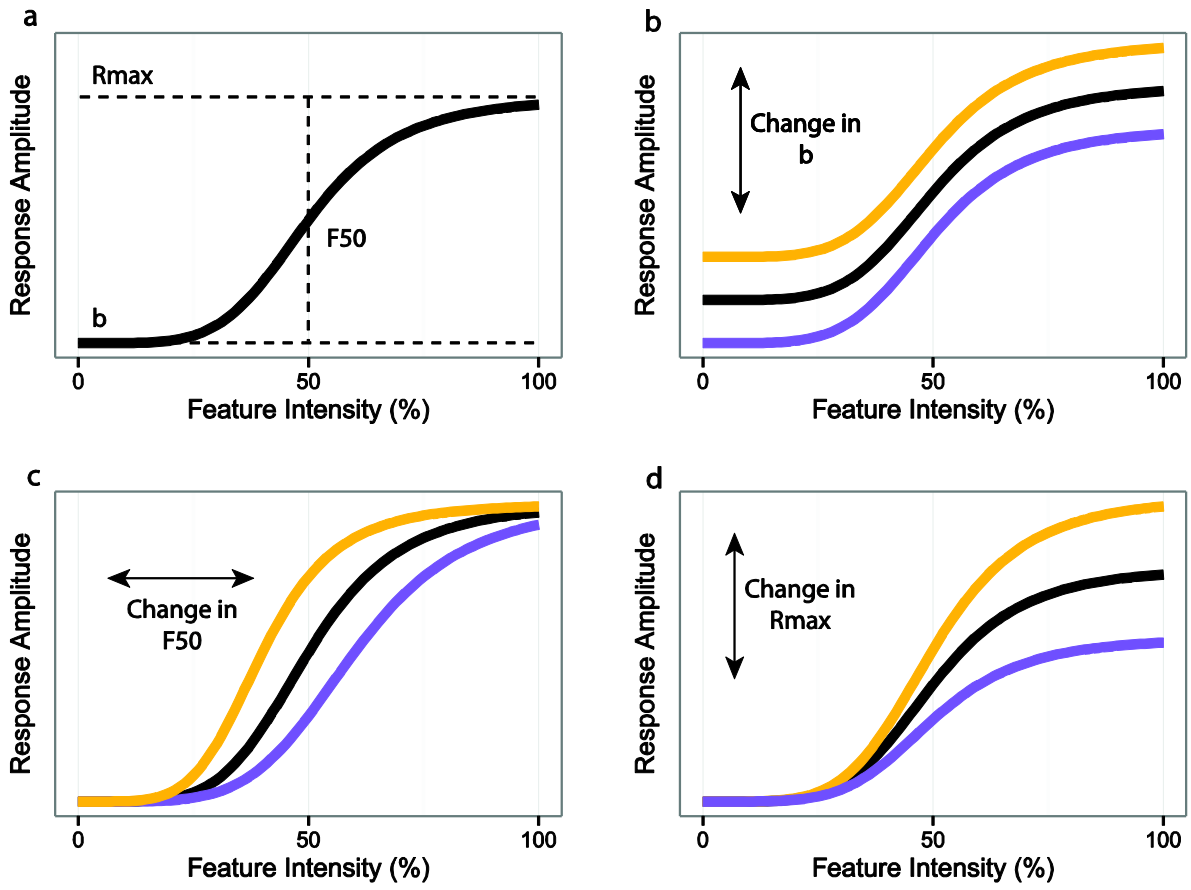


**Figure 6**

Models of Feature Cueing. (a) We modeled the response amplitude to feature-intensity according to the Naka-Rushton equation, using the de-convolved response to each feature (contrast, coherence) and cueing condition (cued, uncued). See text for details. We expected attention to act as a signal modulator in one of four possible ways: baseline shift, feature gain, response gain, or selection. The selection model makes the prediction that a downstream process will be modulated by feature attention, altering how individual feature signals are interpreted by the decision process. (b) Baseline Shift Model. If attention modulates responses by increasing all responses across feature intensities with a constant change, we expect to fit the response gain model to the output. (c) The feature gain model predicts a horizontal shift in the response function; this is equivalent to changing the perceived feature intensity at constant response amplitude. (d) The response gain model predicts that feature intensity will cause a multiplicative gain in response amplitudes during cueing. Figure adapted from Soma et al. (Soma, Shimegi, Suematsu, & Sato, 2013).

**References**


Cohen, M. a, Alvarez, G. a, & Nakayama, K. (2011). Natural-scene perception requires attention. *Psychological Science*, *22*(9), 1165–72. http://doi.org/10.1177/0956797611419168

Cutrone, E. K., Heeger, D. J., & Carrasco, M. (2014). Attention enhances contrast appearance via increased input baseline of neural responses. *Journal of Vision*, *14*(14), 16.

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage*, *9*(2), 179–194.

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *NeuroImage*, *39*(2), 647–660. http://doi.org/10.1016/j.neuroimage.2007.09.034

Feinberg, D. A., & Setsompop, K. (2013). Ultra-fast MRI of the human brain with simultaneous multi-slice imaging. *Journal of Magnetic Resonance*, *229*, 90–100. http://doi.org/10.1016/j.jmr.2013.02.002

Gardner, J. L., Sun, P., Waggoner, R. A., Ueno, K., Tanaka, K., & Cheng, K. (2005). Contrast adaptation and representation in human early visual cortex. *Neuron*, *47*(4), 607–20. http://doi.org/10.1016/j.neuron.2005.07.016

Grill-Spector, K., & Malach, R. (2004). The Human Visual Cortex. *Annual Review of Neuroscience*, *27*(1), 649–677. http://doi.org/10.1146/annurev.neuro.27.070203.144220

Huk, A. C., Dougherty, R. F., & Heeger, D. J. (2002). Retinotopy and functional subdivision of human areas MT and MST. *The Journal of Neuroscience*, *22*(16), 7195–7205.

Itthipuripat, S., Ester, E. F., Deering, S., & Serences, J. T. (2014). Sensory Gain Outperforms Efficient Readout Mechanisms in Predicting Attention-Related Improvements in Behavior. *Journal of Neuroscience*, *34*(40), 13384–13398. http://doi.org/10.1523/JNEUROSCI.2277-14.2014

Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: two distinct brain processes. *Trends in Cognitive Sciences*, *11*(1). http://doi.org/10.1016/j.tics.2006.10.012

Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, *39*(16), 2729–2737.

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(14), 9596–601. http://doi.org/10.1073/pnas.092277599

Mack, A., & Clarke, J. (2012). Gist perception requires attention. *Visual Cognition*, *20*(3), 300–327. http://doi.org/10.1080/13506285.2012.666578

Mack, Arien, & Rock, Irvin. (1998). *Inattentional Blindness.* The MIT Press.

Nestares, O., & Heeger, D. J. (2000). Robust multiresolution alignment of MRI brain volumes. *Magnetic Resonance in Medicine*, *43*(5), 705–715.

Pestilli, F., Carrasco, M., Heeger, D. J., & Gardner, J. L. (2011). Attentional enhancement via selection and pooling of early sensory responses in human visual cortex. *Neuron*, *72*(5), 832–46. http://doi.org/10.1016/j.neuron.2011.09.025

Soma, S., Shimegi, S., Suematsu, N., & Sato, H. (2013). Cholinergic modulation of response gain in the rat primary visual cortex. *Scientific Reports*, *3*. http://doi.org/10.1038/srep01138

Taylor, M., & Creelman, C. D. (1967). PEST: Efficient estimates on probability functions. *The Journal of the Acoustical Society of America*, *41*(4A), 782–787.

Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007). Visual Field Maps in Human Cortex. *Neuron*, *56*(2), 366–383. http://doi.org/10.1016/j.neuron.2007.10.012

Weibull, W. (1951). A Statistical Distribution Function of Wide Applicability. *Journal of Applied Mechanics*.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function I: Fitting, sampling, and goodness of fit. *Perception & Psychophysics*.