**Interactions Between Feature Representations are**

**Well Predicted by a Hierarchical Model of Visual Cortex**

**Or**

**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. Only certain types of attentional modulations may modify the neural activity responsible for scene recognition. For example, attending to motion appears to result in diminished awareness of scene gist (Cohen, Alvarez, & Nakayama, 2011), while attending to other features may not cause inattentional blindness. We propose that this feature-specific hypothesis may depend crucially on the layout of visual cortex. Our hypothesis is that modifying a neural representation through spatial or feature-based attention is only detrimental to awareness of other features when there are direct connections between the relevant neural representations. Contrast is well specified by neural activity in the earliest cortical visual area, V1. The human MT+ complex on the other hand is less sensitive to contrast, but considerably more sensitive to the presence or absence of motion in a stimulus. We propose to test our hypothesis by measuring behavioral and neural activity related to the representations of contrast and motion when participants are cued to a feature. Our hypothesis predicts that attention to contrast will be detrimental to the perception of motion, but not vice versa.

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 attention acts as a form of either response gain (Itthipuripat, Ester, Deering, & Serences, 2014) or a kind of selective gating of activations (Pestilli, Carrasco, Heeger, & Gardner, 2011). But little is known about how attention impacts downstream processes. Attentional load theory (see e.g. Lavie, 2014) for example predicts that ‘perceptual load’ acts as a gating function to awareness. When load is high, only attended features will reach awareness, and vice versa. Our hypothesis here is in line with this general architecture. In line with the neurophysiological evidence of attention as a neural modulator we believe that the gating of awareness can be understood as a function of interactions between neural populations. Although attention may not explicitly modify “unattended” neural representations there may nevertheless be cross-talk due to reciprocal connections across cortex. In visual cortex, where the majority of connections are feed-forward in a hierarchical structure, this hypothesis makes a specific and testable prediction: attention to hierarchically ‘higher’ features should not impact the awareness of hierarchically ‘lower’ features, to the extent that attention is limited to modulating specific neural representations. We approached this problem by designing an experiment in which we could measure, behaviorally, the effect of attention on unattended stimuli. Simultaneously we collected functional magnetic resonance imaging (fMRI) measurements of the blood-oxygenation level dependent (BOLD) response. There is evidence that for low-level features such as contrast the BOLD response captures the entire variance in behavioral performance on simple discrimination tasks (Hara & Gardner, 2014; Pestilli et al., 2011). By recording both of these measurements we can therefore set up a computational model that predicts the magnitude of behavioral change expected for any change in neural response. Not only do we expect asymmetrical relationship in behavior, but we expect to be able to explain the variance in this relationship according to a similar asymmetrical relationship in the neural responses.

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). \*\*\*TODO\*\*\*

We view attention in visual space as a local context-dependent modulation of neural activity in precise regions of visual cortex. In contrast, visual awareness appears to be a downstream property of neural activation that is dependent on activations in visual cortex. Based on these relatively well-established assumptions we predict that any asymmetries in how attention interacts with awareness should be well predicted by the topology and interconnections in cortex. For example, the unexpected result that scene gist is abolished by motion (Cohen et al., 2011) but not by other demanding attentional tasks (Li et al., 2002) may be explained by the interconnections between the regions that underlie scene recognition and those that are responsible for representing motion. Based on this hypothesis we predicted that in a task where contrast and motion coherence are set up, as opposing features performance would be asymmetrically reduced on motion judgments when participants are cued to attend to contrast. We expect that this relationship will be precisely explained by the variability, due to attention, in neural responses captured by the BOLD response 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. An architectural constraint such as the one we have outlined here would have impacts not only on visual awareness but also on memory, cognitive control, learning, and on any other downstream processes that the brain computes.

**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. Subjects in the behavioral experiment performed one training session 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). Subjects in the functional experiment performed a retinotopic mapping (1 hr, consisting of ten 4 minute scans) and four to five sessions of the main experiment (2 hrs each, consisting of ten 7 minute scans).

**Experimental Task**

Subjects performed a two-alternative forced choice contrast or motion discrimination task. On each trial participants were shown two patches of dots and asked to report which had a higher apparent contrast or motion coherence. Participants were cued on each run 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 random checkerboards at 55% contrast at 40hz. A random inter stimulus interval followed for between 200 and 500 ms after which the fixation cross turned white, indicating that the response period was starting. Participants had 1 second to respond. Each trial was followed by a random inter-trial-interval of 300 to 500 ms. Participants fixated a central cross (1 deg x 1 deg visual angle, 1 pixel wide, luminance ?!). During control runs participants were cued to attend to either motion or contrast and responded about the cued feature. On task runs participants were informed after stimulus presentation that they should respond about the miscued feature. These task miscued trials occurred 15% of the time and were interspersed pseudo-randomly. The trial timing was modified during scanning to improve estimation of the hemodynamic response, stimulus: 750 ms, mask: 250 ms, ISI: 200-1000 ms, resp: 1000 ms, ITI: 2000-10000 ms. During the behavioral experiment the screen updated at 100 hz and at 60 hz during scanning.

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 equally brighter. Each dot was 4 x 4 pixels on the screen. The luminance difference between the dots and the background was defined as 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. All dots moved at a consistent speed of 3.25 degrees / s.

Contrast and motion discrimination performance was tested at a single pedestal intensity. Contrast was tested at 60% and motion coherence was tested at 10%. Both features were crossed such that neither feature was informative about the strength of the other feature on any given trial. For each feature a 1-up-3-down staircase (??) was used to set the increments in contrast or motion coherence that was 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 a near-threshold level, eliminating any potential confound with task difficulty between conditions. 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. For each pedestal an independent staircase was computed to maintain task difficulty near threshold. 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 (?) cm from the subject’s eyes to obtain a field of view of ? x ?. 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. The projector and LCD screen were calibrated to have linearized gamma scales using a PR650 Spectroradiometer (Photo Research Inc., Chatsworth, CA.). 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>).

**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 all sessions. 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.

Todo: eye position analysis

**Contrast and Motion Discrimination Functions**

Feature discrimination task performance was evaluated using feature-discrimination functions. 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 (c ) = ….

Where p(delta i) is the probability of being correct given an intensity increment of delta I, lambda is the lapse rate, epsilon 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 X psychometric functions with X 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 were able to compute the variability of these functions across runs.

For each participant we estimated their discrimination functions for control runs, task cued trials, and task miscued trials.

**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 ??, TI ??, TE ??, FA ??, voxel size ??, matrix ??. 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 ??, voxel size 2.5 x 2.5 x 2.5 mm, matrix 88 x 88). An automated procedure was used to find the best affine transform to align the session and canonical anatomy (Nestares & Heeger, 2000). The functional scans were aligned to the session anatomy directly using the coordinates measured by the scanner. Retinotopic mapping was performed using a T2\*-weighted functional scan (multiband 2, TR 1400 ms, TE 30 ms, flip angle ??, 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 images were analyzed through a pipeline using mrTools.

**Retinotopy**

Visual fields were determined based on a retinotopy performed 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, ??) was fit to each voxel, identifying the Gaussian response field parameters that best fit the recorded response data. Visual fields were then defined according to published criteria.

**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. 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, Dumoulin, & Brewer, 2007), as well as with our functional localizer and found significant overlap between voxels identified with each procedure. See Figure X for additional information.

**Feature-response Functions**

To compute the feature-response functions, a deconvolution analysis (for details see: (Gardner et al., 2005)) was used to determine the mean hemodynamic response to each dot patch in the contralateral visual cortex. 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 fit to the Naka-Rushton equation

**Results**

Participants in our experiment performed a two-alternative forced 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 run. Each participant 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 180 minutes to learn and perform the coherence discrimination task successfully. Subjects then completed as many runs of type 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.

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 plot. We chose to normalize the results by the performance in the control runs to better estimate

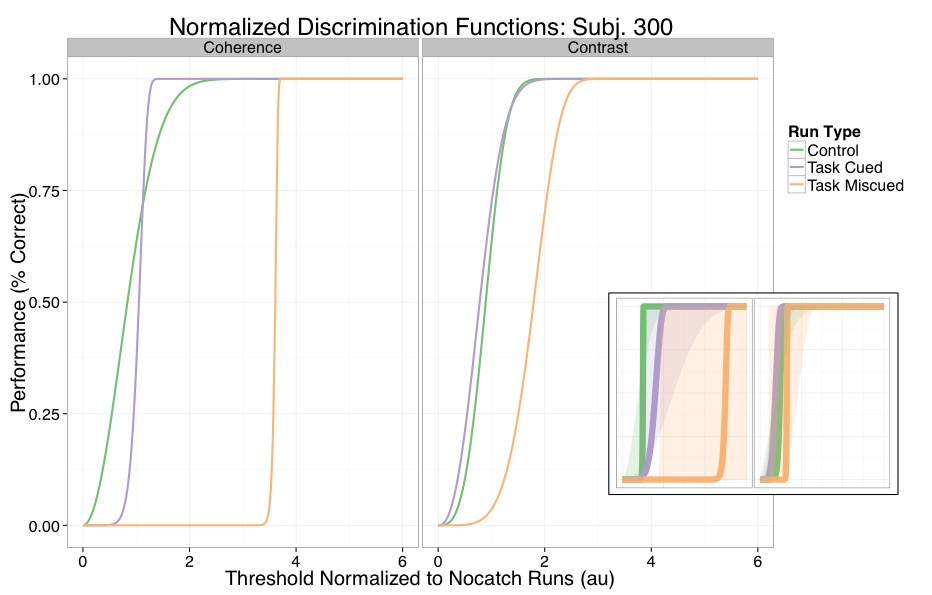
**Discussion**

Our hypothesis is that the hierarchical organization in visual cortex is sufficient to predict the asymmetrical interactions that can occur between features in visual tasks. For example, it is well known that under demanding attention conditions our ability to report the ‘gist’ of a scene remains largely unimpaired (Li et al., 2002). But this effect breaks down under specific experimental conditions, for example when the attention demanding task requires motion tracking (Cohen et al., 2011, p. 201). We propose that reconciling these results requires an understanding of how neural activity in visual cortex is interconnected. In our design we asked participants to report responses about a single feature, either stimulus contrast or motion coherence, while simultaneously manipulating the intensities of both features. This allowed us to estimate the impact of being cued and asked to respond about different features (task: miscued), compared to responding about the cued feature. We found behavioral results that supported our hypothesis, showing that miscuing coherence is more detrimental to performance than miscuing contrast. We recorded BOLD responses in visual cortex to investigate whether specific voxel populations might be predictive of this asymmetrical cueing relationship. We found that ??!? I don’t know!?!?

rainbowRaid:Users:dan:proj:att_awe:analysis:figures:behav:multi_plot.pdf

**Figure X**

Estimated Discrimination Thresholds Across Participants

****

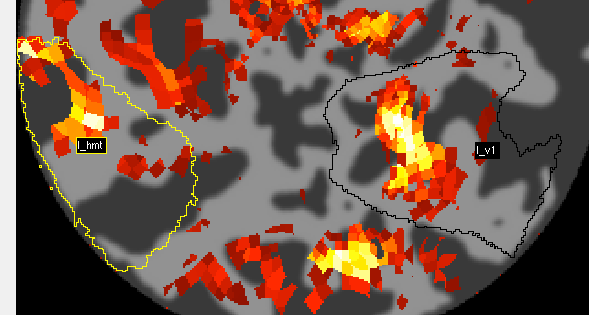
**Figure Z**

Normalized discrimination functions, averaged across participants.

**rainbowRaid:Users:dan:proj:att_awe:analysis:figures:roi:left_hemisphere_ROIs.pdf**

**Figure Y**

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 voxel activations to an R^2 value of greater than 0.1 and excluding voxels in the retinotopically defined early visual areas (V1, V2, and V3).

****

**Figure W**

Event-Related Deconvolutions 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 anatomically defined V1 and hMT+ ROIs in a left hemisphere occipital cortex flat map. Heatmap 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 belowt he midline.

**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

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

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

Hara, Y., & Gardner, J. L. (2014). Encoding of graded changes in spatial specificity of prior cues in human visual cortex. *Journal of Neurophysiology*. http://doi.org/10.1152/jn.00729.2013

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.

Lavie, N. (2014). Blinded by the load: attention, awareness and the role of perceptual load. *… of the Royal …*, (March). Retrieved from http://rstb.royalsocietypublishing.org/content/369/1641/20130205.short

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, 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

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*.