Modeling peripheral visual acuity enables discovery of gaze strategies at multiple time scales during natural scene search

Pavan Ramkumar
Department of Physical Medicine and Rehabilitation, Northwestern University and Rehabilitation Institute of Chicago, Chicago, IL, USA
Department of Neurobiology, Northwestern University, Evanston, IL, USA

Hugo Fernandes
Department of Physical Medicine and Rehabilitation, Northwestern University and Rehabilitation Institute of Chicago, Chicago, IL, USA
Instituto Gulbenkian de Ciência, Oeiras, Portugal

Konrad Kording
Department of Physical Medicine and Rehabilitation, Northwestern University and Rehabilitation Institute of Chicago, Chicago, IL, USA

Mark Segraves
Department of Neurobiology, Northwestern University, Evanston, IL, USA

Like humans, monkeys make saccades nearly three times a second. To understand the factors guiding this frequent decision, computational models of vision attempt to predict fixation locations using bottom-up visual features and top-down goals. How do the relative influences of these factors evolve over multiple time scales? Here we analyzed visual features at fixations using a retinal transform that provides realistic visual acuity by suitably degrading visual information in the periphery. In a task in which monkeys searched for a Gabor target in natural scenes, we characterized the relative importance of bottom-up and task-relevant influences by decoding fixated from nonfixated image patches based on visual features. At fast time scales, we found that search strategies can vary over the course of a single trial, with locations of higher saliency, target-similarity, edge–energy, and orientedness looked at later on in the trial. At slow time scales, we found that search strategies can be refined over several weeks of practice, and the influence of target orientation was significant only in the latter of two search tasks. Critically, these results were not observed without applying the retinal transform. Our results suggest that saccade-guidance strategies become apparent only when models take into account degraded visual representation in the periphery.

Introduction

During naturalistic visual search or exploration, in order to bring objects of interest to the fovea, we move our eyes about three times a second. What factors influence where we look? The dominant working hypothesis for computational models of gaze suggests that the visual system computes a priority map of the visual field before every saccade, and a saccade is subsequently made to a target of high priority (Bisley & Goldberg, 2010; Fecteau & Munoz, 2006; Serences & Yantis, 2006; Treisman & Gelade, 1980; Wolfe, 1994).

What factors constitute this priority map? Studies have shown that many factors influence the guidance of eye movements. These factors can be generally grouped into task-independent, or *bottom-up* features, and *task-relevant* features. Examples of bottom-up features are luminance contrast (Reinagel & Zador, 1999; but see Einhäuser & König, 2003), color contrast (Itti, Koch, & Niebur, 1998), energy (Ganguli, Freeman, Rajashekar, & Simoncelli, 2010), and saliency (Itti & Koch, 2001; Itti et al., 1998); examples of task-specific features are relevance (resemblance to search target; Einhäuser, Rutishauser, & Koch, 2008); learned context about

Citation: Ramkumar, P., Fernandes, H., Kording, K., & Segraves, M. (2015). Modeling peripheral visual acuity enables discovery of gaze strategies at multiple time scales during natural scene search. *Journal of Vision, 15*(3):19, 1–20. http://www.journalofvision.org/content/15/3/19, doi: 10.1167/15.3.19.

doi: 10.1167/15.3.19 Received June 4, 2014; published March 26, 2015 ISSN 1534-7362

target location (Ehinger, Hidalgo-Sotelo, & Torralba, 2009); factors that are ecologically relevant but not specific to the task, such as intrinsic value (Gottlieb, 2012); and exploratory strategies for minimizing uncertainty (Gottlieb, Oudeyer, Lopes, & Baranes, 2013; Renninger, Verghese, & Coughlan, 2007). In addition, change in direction between successive saccades (Wilming, Harst, Schmidt, & König, 2013) and natural statistics of saccade magnitude and direction (Tatler & Vincent, 2009) also enable saccade prediction. Predictive models of eye movements are derived from priority maps comprising one or more of the above factors.

Although systematic comparisons of factors influencing saccade guidance have been made in humans (e.g., Hwang, Higgins, & Pomplun, 2009; Navalpakkam & Itti, 2006), similar studies are few and far between for nonhuman primates (NHPs; Berg, Boehnke, Marino, Munoz, & Itti, 2009; Fernandes, Stevenson, Phillips, Segraves, & Kording, 2013; Kano & Tomonaga, 2009). Since NHPs are important model organisms for understanding complex tasks performed by the visual system, it is important to rigorously model their gaze behavior in naturalistic conditions. However, although a multitude of factors constituting priority maps have been proposed in different contexts, how this notion of priority in the visual field evolves over time has not been addressed. Therefore, our goal in this study is to understand how search strategy in monkeys evolves over multiple time scales in naturalistic conditions, ranging from a few seconds of exploring a single image, to learning to perform similar search tasks over several weeks.

To this end, we designed a natural search task in which two macaque monkeys searched for either a vertical or a horizontal Gabor target placed randomly within human-photographed scenes. To analyze gaze behavior from this task, we adopted two critical methodological innovations. First, although it is well known that visual features in the periphery are not sensed with the same acuity as those at the fovea, very few models of gaze have taken this into account (Zelinsky, 2008). Therefore, before computing visual features at fixation, we applied a simple computational retinal transform centered on the previous fixation that degrades peripheral visual information in a retinally realistic manner (Geisler & Perry, 1998). Second, we note that the features that comprise priority maps are often correlated; for instance, edges in natural scenes tend to have both high luminance contrast and high energy, and the apparent influence of luminance contrast might potentially be explained away by the influence of energy, or vice versa. Therefore, to tell apart the relative influence of correlated features, we applied multivariate decoding analysis of fixated versus nonfixated patches to quantify the influence of various visual features on gaze.

Over the course of viewing a single image, we found that monkeys fixated locations of greater saliency, target similarity (relevance), edge-energy, orientedness, and verticalness later on in the trial. Monkeys used short saccades to select locations scoring high in all features with the exception of orientedness, which was selected for using long saccades. We also found a significant practice effect over several weeks, with target orientation having a significant effect only in the latter of two tasks performed. Thus, realistic modeling of peripheral vision and multivariate decoding allowed us to tease apart the relative influence of various features contributing to the priority map for saccade guidance and to track their influence over multiple time scales of interest.

Methods

Animals and surgery

Two female adult monkeys (*Macaca mulatta*) aged 14 and 15 years, and identified as MAS15 and MAS16, participated in these experiments. MAS15 received an aseptic surgery to implant a subconjunctival wire search coil to record eye movements. Eye movements of MAS16 were measured using an infrared eye tracker (ISCAN Inc., Woburn, MA; http://www.iscaninc.com/).

Stimuli and eye-movement recordings

Monkeys viewed grayscale human-photographed natural scenes (Fernandes et al., 2013; Phillips & Segraves, 2010). Each scene included a 2° × 2° Gabor wavelet target with position chosen randomly across a uniform distribution within the scene (Figure 1). The Gabor target was alpha-blended with the underlying scene pixels giving it a transparency of 50%. Scenes were chosen from a library of 575 images, in a pseudorandom sequence and each scene was presented in, at most, 10 trials with the Gabor target placed in different locations for each of those 10 trials.

In each trial, monkeys were given at most 20 saccades to find the target. Once the target was found, they had to fixate for a minimum of 300 ms to receive a water reward of approximately 0.20–0.25 ml per trial. Both monkeys were highly familiar with searching in natural scenes and had done versions of our task for previous studies earlier.

We used the REX system (Hays, Richmond, & Optican, 1982) based on a PC computer running QNX (QNX Software Systems, Ottawa, Ontario, Canada), a

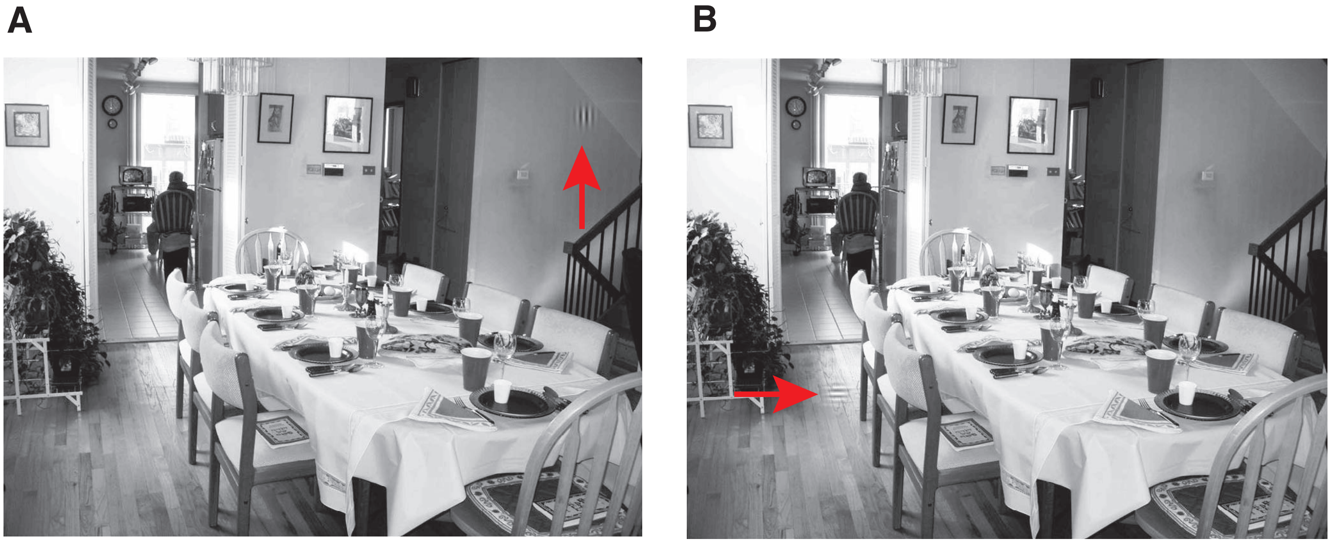

Figure 1. Experimental design. Two monkeys were asked to search for 2° × 2° Gabor-wavelet targets of either (A) vertical or (B) horizontal orientation, alpha-blended into natural scenes. Red arrows (not displayed to the monkeys) indicate the locations of the target. These target locations were distributed randomly with a uniform distribution across the scene. For each trial, the monkeys were permitted, at most, 20 fixations to locate the target.

real-time UNIX operating system, for behavioral control and eye position monitoring. Visual stimuli were generated by a second, independent graphics process (QNX–Photon) running on the same PC and rear-projected onto a tangent screen in front of the monkey by a CRT video projector (Sony VPH-D50, 75 Hz noninterlaced vertical scan rate, 1024 × 768 resolution). The distance between the front of the monkey's eye and the screen was 109 cm. The stimuli spanned 48° × 36° of visual field.

Saccade detection

Saccade beginnings and endings were detected using an in-house algorithm. We used thresholds of 100°/s for start and stop velocities and marked a saccade starting time when the velocity increased above this threshold continuously for 10 ms. Likewise, saccade-ending times were marked when the velocity decreased continuously for 10 ms and fell below 100°/s at the end of this period of decrease. We considered saccades longer than 150 ms as potentially resulting from eye-blinks or other artifacts and discarded them. Fixation locations were computed as the median (x, y) gaze coordinate in the intersaccadic interval. Since monkeys are head-fixed, they tend to make a few long-range saccades that are unlikely to occur in a naturalistic environment where the head is free. To mitigate the potential effect of large saccades, we restricted our analysis to saccades under 20° (we also reanalyzed the data with all saccades and did not find any qualitative differences in our results).

Modeling peripheral visual acuity with a retinal transform

It is well known that visual features cannot be resolved in the periphery with the same resolution as at the point of gaze and that the representation of visual information in the brain is retinotopic. Despite this, the most influential models of eye gaze use visual features at fixation at full resolution, and most bottom-up saliency maps are computed in image-centered coordinates (e.g., Itti et al., 1998). However, when the oculomotor system computes a saliency map for each saccade, it can only work with retinotopic representations of local image statistics in the visual cortex. Thus, models of priority that take into account degraded representations of peripheral information (e.g., Hooge & Erkelens, 1999; Zelinsky et al., 2013) and analysis methods that look for effects of visual features at different saccade lengths might provide insight into the phenomena underlying the computation of priority in the brain. To model the peripheral sensitivity of the retina in a more realistic manner, we implemented a simple blurring filter.

The degraded image can be obtained as a parametric blurring of the original image, with the spatial extent of the blur filter scaled by the distance from the previous fixation. In practice, we precomputed a Gaussian pyramid filter bank at three levels and convolved it with the original image. We then selected appropriate pyramids for each eccentricity according to the implementation described in Geisler and Perry (1998).

First, we transformed the natural image to a Gaussian pyramid. An image pyramid is a collection of trans-

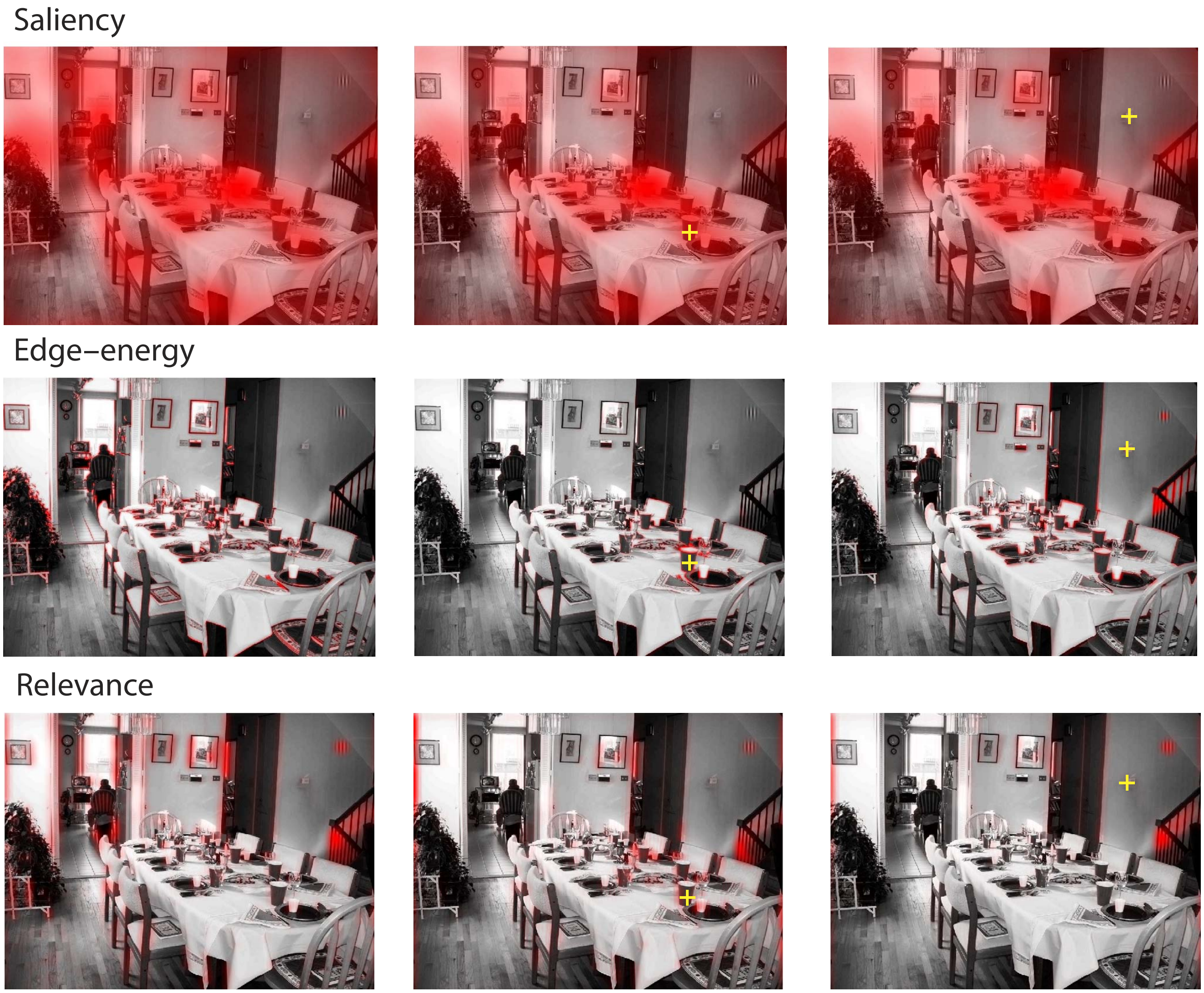

Figure 2. The effect of peripheral blurring on sensitivity to visual features. Heat maps of saliency, edge–energy and relevance are overlaid on the retinally transformed images. Left: No blur filter. Center and right: Blur filter applied at two different fixation locations indicated by the yellow crosshair.

formed images, each typically obtained by applying a convolution to the original image. Here, we used a 3-level Gaussian pyramid, with each level downsampled by a factor of 2 and convolved with a Gaussian filter (Supplementary Figure S1). We used the implementation provided by the Matlab Pyramid Toolbox (http://www.cns.nyu.edu/lcv/software.php). For each pyramid level, we calculated the maximally resolvable spatial frequency *f* as the Nyquist rate (half the sampling rate).

Next, using the method by Geisler and Perry (1998), we calculated the critical eccentricity for a given resolvable spatial frequency, according to known retinal physiology. The critical eccentricity e_c is given by:

$$e_c = \frac{e_2}{\alpha f} \ln\left(\frac{1}{CT_0}\right) - e_2 \qquad (1)$$

where e_2 is the half-resolution eccentricity, α is the spatial frequency decay constant, f is the Nyquist rate, and CT_0 is the minimum contrast threshold. We used the standard parameters of $e_2 = 2.3$, $\alpha = 0.106$, and $CT_0 = 0.04$ used by Geisler and Perry (1998).

We then swept through the image pixel-by-pixel and assigned the grayscale value of the pyramid level whose best-resolvable frequency exceeds the one that can be resolved at the eccentricity of that pixel with respect to a given fixation location. Once this retinal transform is applied, the detectability of high-frequency features from far away is clearly degraded (Figure 2).

Computation of visual features at fixation

To understand visual features that drive gaze behavior, we computed a number of features from image

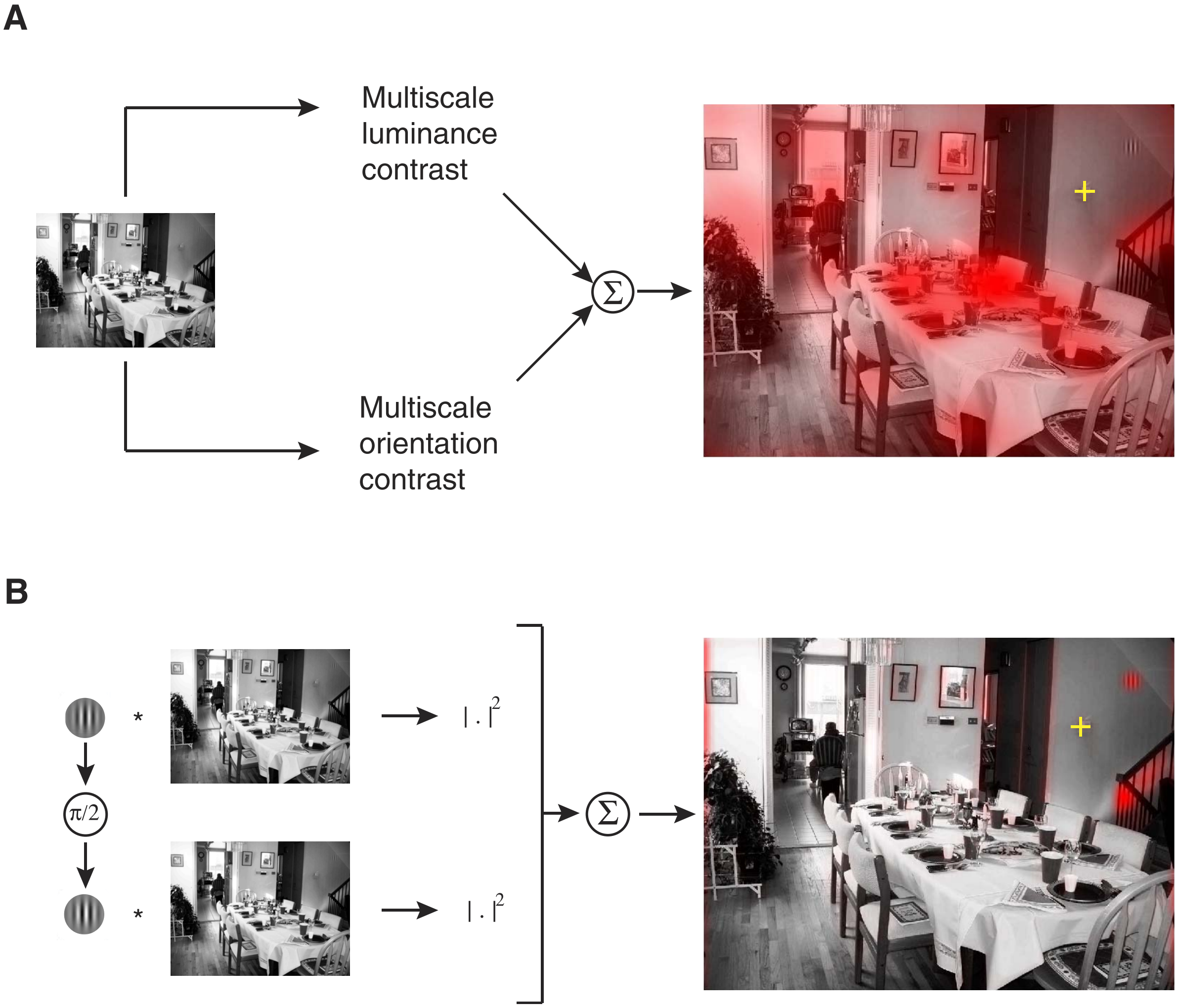
A
Multiscale luminance contrast
Multiscale orientation contrast
Σ
B
*
π/2
*
| . |²
| . |²
Σ

Figure 3. Computation of priority maps based on bottom-up saliency and top-down relevance. The yellow crosshair represents the fixation with respect to which the peripheral blurring operator is applied. (A) Saliency: Itti-Koch saliency (Itti et al., 1998) computes luminance contrast and orientation contrast at four different scales and eight different orientations, and adds them across scales to produce a saliency map. (B) Relevance: A Gabor-detector computes template Gabor filter outputs for a quadrature pair (in-phase Gabor and its 90° phase shift) and takes the sum of squares to produce a relevance map. Vertical and horizontal relevance maps are computed for modeling the vertical and horizontal tasks, respectively. Only the vertical relevance map is shown here.

patches at fixated locations after peripherally blurring them using a retinal transform. The patch size specifies the size of the spotlight in the periphery over which visual features are pooled prior to saccade decisions. We chose patch sizes of 1° × 1°, 2° × 2°, and 4° × 4° and performed all analyses separately for each patch size.

We computed one bottom-up feature (saliency), one task-specific feature (relevance), and three low-level orientation statistics (edge–energy, orientedness, and predominant orientation).

Bottom-up saliency

Bottom-up saliency is the extent to which an image patch stands out with respect to its immediate neighborhood. Typically, saliency at a particular location is operationally defined as the difference between a feature's (or set of features) value at that location and the feature's (or set of features) value in the immediately surrounding neighborhood. This center-surround difference map is typically computed for multiple features at multiple scales and then summed together to obtain a saliency map (Itti & Koch, 2001). A number of bottom-up saliency maps have been proposed (see Borji, Sihite, & Itti, 2013 for a survey). We used the most popular model (Itti et al., 1998) implemented in the Matlab Saliency Toolbox (www.saliencytoolbox.net; Walther & Koch 2006) with eight orientations and four scales (see Figure 3A).

Top-down relevance or target-similarity

The relevance of a fixated image patch must measure the extent to which the information in the patch is relevant for the search task. We assume a greedy searcher who looks at patches that maximally resembled the target.

Before computing a similarity measure to the target, we locally standardized the grayscale values of the patch and then linearly stretched them to lie between 0 and 1.

$$Z(x,y) = \frac{I(x,y) - \mu}{\sigma} \tag{2}$$

$$U(x,y) = \frac{Z(x,y) - \min(Z)}{\max(Z) - \min(Z)} \tag{3}$$

We then defined target relevance of a fixated image patch as the sum of squares of the convolution of the fixated image patch with the target Gabor wavelet and its quadrature phase pair:

$$R = (U^*G)^2 + (U^*G')^2 \tag{4}$$

where G and G' are the Gabor target and its 90° phase-shifted version and * represents two-dimensional convolution. The scalar relevance value that is used for subsequent analysis is then the central pixel value of the convolved image, which happens to be equivalent to the sum of squares of dot products between the patch and the quadrature Gabor pairs for the 2° × 2° image patch.

For illustration, the relevance maps visualized in Figures 2 and 3B are obtained by taking the sum of squares of the convolution between the target Gabor quadrature pairs and the full-sized contrast-normalized image.

Orientation statistics

Aside from the composite measures of saliency and relevance, we extracted local orientation statistics using an eigenvalue decomposition of the energy tensor computed as the covariance matrix of horizontal and vertical edge gradients at each pixel (Figure 4; Ganguli et al., 2010):

$$I_x(x,y) = \frac{\partial I(x,y)}{\partial x} \tag{5}$$

$$I_y(x,y) = \frac{\partial I(x,y)}{\partial y} \tag{6}$$

$$C(x,y) = \begin{bmatrix} cov(I_x, I_x) \\ cov(I_x, I_y) \end{bmatrix} \begin{bmatrix} cov(I_x, I_y) \\ cov(I_y, I_y) \end{bmatrix} \tag{7}$$

where I_x and I_y are horizontal and vertical gradients, and $C(x, y)$ is the energy tensor. Horizontal and vertical image gradients were computed as differences between consecutive columns and rows of the image, respectively. From the eigenvalues of $C(x, y)$, λ_1, and λ_2, we then computed the following orientation statistics.

The edge–energy of the patch is given by $E = \lambda_1 + \lambda_2$, the eigendirection, $\theta = \tan^{-1}\{[cov(\mathrm{I}_x, \mathrm{I}_x) - \lambda_2]/[cov(\mathrm{I}_x, \mathrm{I}_y)]\}$ gives the *predominant orientation*, and the ratio of variance along the eigendirection and its orthogonal direction, $\zeta = (\lambda_1 - \lambda_2)/(\lambda_1 + \lambda_2)$ gives the *orientedness* of the patch (0 for no orientation, 1 for bars and edges in any direction). We then computed measure for *verticalness* as $\upsilon = |\sin \theta|$. Verticalness is 1 for a perfectly vertical patch and 0 for a perfectly horizontal patch.

Analysis of fixation statistics

Our goal was to quantify the relative importance of each visual statistic in predicting fixation behavior and to study how these influences evolved over different time scales. The typical method to measure the predictive power of a visual statistic is to compare its value at fixated versus nonfixated points in the stimulus image. However, the stimuli are human-photographed scenes, with objects of interest near the center; therefore, fixated locations tend to be closer to the center of the image rather than randomly sampled nonfixated locations. This is known as the center bias, and can potentially inflate effect size for a given statistic (e.g., Kanan, Tong, Zhang, & Cottrell, 2009; Tseng, Carmi, Cameron, Munoz, & Itti, 2009). To avoid center bias, shuffled-control statistics are taken from the same location as the current fixation, but from a different image. In all subsequent analyses of visual statistics at fixation, we compared real visual features at fixation with shuffled controls computed as described above, and always excluded the very last fixation to avoid confounds from the target itself.

Fourier analysis

Local power spectra of natural image patches are highly informative about scene texture and structure (Oliva & Torralba, 2001) and can be diagnostic of natural scene categories (Torralba & Oliva, 2003). In particular, the two-dimensional Fast Fourier Transform (FFT) allows us to visualize Fourier amplitude at different spatial frequencies and orientations at the same time. To obtain a qualitative understanding of whether visual statistics at fixation were biased by the sought-after Gabor target, we studied the similarity between the target and fixated image patches by examining their spectral properties (see e.g., Hwang et al., 2009). We computed local Fourier amplitude (absolute value of the FFT) in 1° × 1°, 2° × 2°, and 4° × 4° windows at each fixation using the real image and a

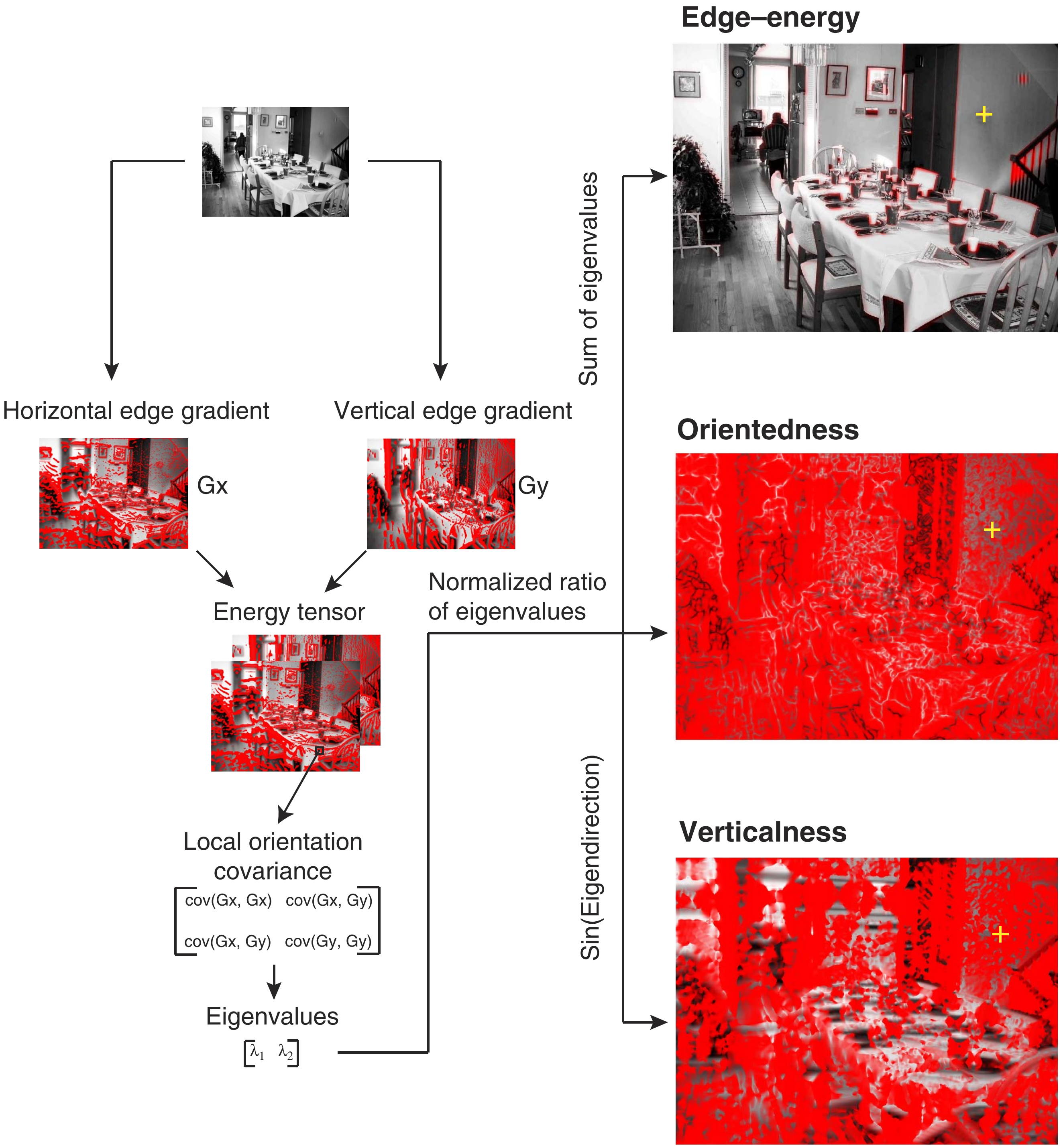

Figure 4. Orientation statistics. For each natural scene, we computed the local orientation statistics at all fixated and shuffled-control patches: edge–energy, defined as the sum of squares of local gradients; *orientedness*, defined as a normalized ratio of eigenvalues of the local orientation covariance matrix, which measures the extent to which a single orientation is predominant; and the *predominant orientation*, defined as the eigendirection. *Verticalness* is defined as the absolute sine of the predominant orientation.

shuffle-control image. We averaged the obtained power spectra across fixations for each trial, and then computed a two-tailed t statistic across trials of the difference in log absolute FFT between fixated patches and shuffled-controls.

Receiver operating characteristic analysis

To study the extent to which each visual feature could predict gaze, we performed a receiver operating characteristic (ROC) curve analysis comparing the distributions of fixated patches and shuffled controls over each visual feature. We computed the area under the ROC curve (AUC) for each feature and bootstrapped across fixations to obtain 95% confidence intervals (CIs) for the AUCs.

Analysis of visual feature influence within a trial

Although analysis of visual features at fixations across trials and sessions provide a general idea of which features influence saccade planning, they implicitly assume that the strategy of planning saccades remains unchanged over time. Since our goal is to obtain an insight into how search strategy evolves over time, we studied the difference between visual features at early and late fixations. To understand how search strategy evolves within a single trial, we computed the Spearman's rank correlation coefficient between different visual features and two eye-movement variables: the numbered order of fixation within a trial (fixation order) and the saccade length made to reach that fixation location, along with their corresponding bootstrapped 95% confidence intervals.

Analysis of visual feature analysis over sessions using multivariate decoding

Although ROC analysis provides a useful characterization of the influence of individual features, it is a metric that compares two one-dimensional distributions. The AUC is a good measure of the importance of a given variable but comparing two AUCs obtained with two different variables is not meaningful as a metric of relative importance if the variables are correlated. However, most features that inform bottom-up saliency are correlated with each other and with local contrast (Kienzle, Franz, Schölkopf, & Wichmann, 2009). Further, in our task, the Gabor-similarity metric (relevance) may be correlated with orientation or edge–energy. Therefore, a large AUC for relevance could merely result from the effect of a large AUC for edge–energy, or vice versa.

Multivariate analysis is a suitable tool for potentially explaining away the effects of correlated variables. For instance, over the past few years, generalized linear models (GLMs; Nelder & Wedderburn, 1972) have been used to model spike trains when neural activity may depend on multiple potentially correlated variables (Fernandes et al., 2013). Here we used a particular GLM, multivariate logistic regression, to discriminate fixated patches from shuffled controls based on the local visual features defined above: saliency, relevance, edge–energy, orientedness, and verticalness, defined as the sine of the predominant orientation.

Using this decoding technique, we estimated effect sizes for each feature using partial models leaving each feature out of the full model comprising all features, and comparing these partial models with the full model (see Appendix A for details on logistic regression, goodness of fit characterization, and effect size estimation). For this analysis, we pooled fixations across four consecutive sessions for each search task and each animal.

Results

To understand the relative contributions of visual features that guide saccades during naturalistic vision we had two monkeys search for embedded Gabor wavelets in human-photographed natural scenes. Monkeys performed the task over two sets of four consecutive days, with the sets differing on the orientation of the target: vertical in one set and horizontal in the other. We modeled the effects of decreasing retinal acuity as a function of eccentricity and analyzed visual features at fixated locations in natural scenes to ask how the features of the target stimulus (task-relevant) as well as other features of the visual scene (bottom-up) affect the search behavior. We then studied how the influence of these visual features evolved across multiple time scales.

Search performance

We first wanted to see if the monkeys could successfully perform the task; that is, if they could successfully find the target in a considerable number of the 700–1,500 trials they completed on each of the eight days. We found that while one monkey (MAS15) performed better than the other (MAS16) as measured by the fraction of successful trials (65% vs. 41%) and by the mean number of fixations required to find the target, (7.1 vs. 8.3), both monkeys were able to find the target in each session (Supplementary Figure S2). Although there are differences in performance, consistent with previously observed behavior of both monkeys (Kuo, Kording, & Segraves, 2012), they were able to perform the task.

Fourier analysis reveals the effect of edge–energy and target orientation

Gabor wavelets have a well-defined spectral signature (Figure 5, first row). Thus, we may reasonably expect saccade target selection to be guided by the local frequency content in natural scenes. To get an insight into this possibility, we computed 2-D Fourier power spectra at fixated and shuffled-control patches

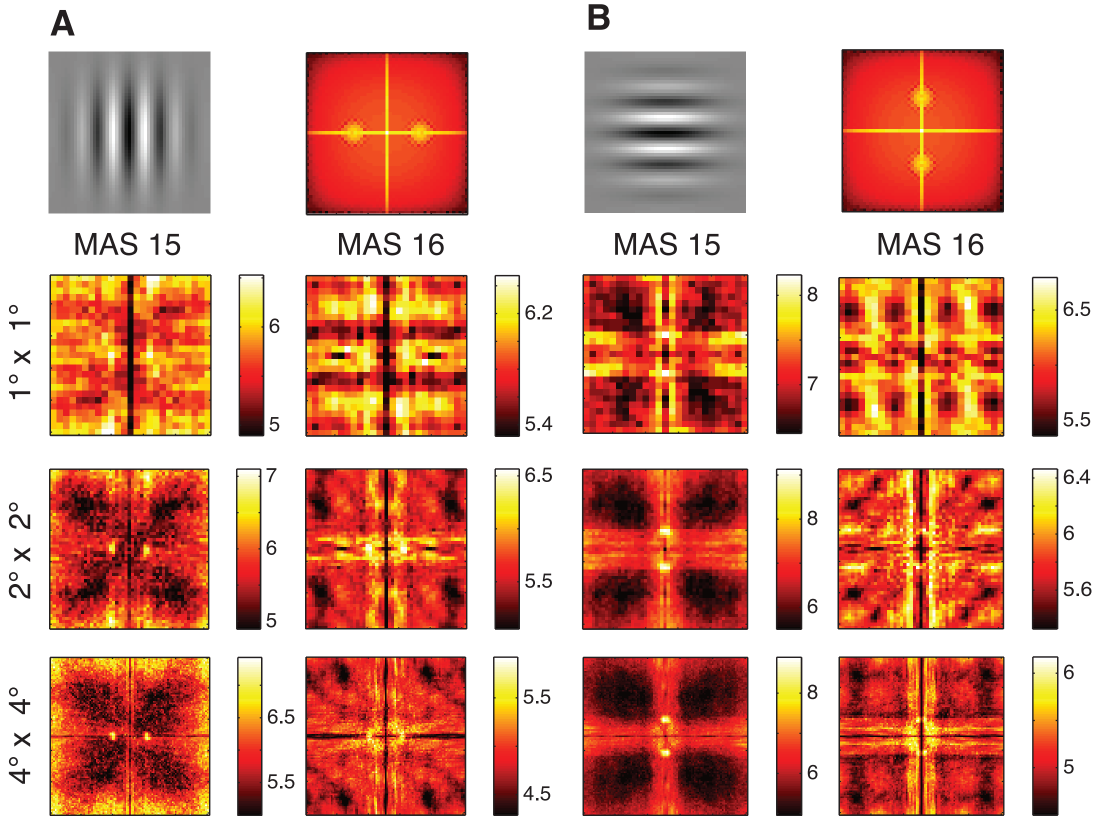

Figure 5. Spectral analysis of fixation locations. Monkey fixation locations exhibit target-dependent orientation bias. A: vertical task, B: horizontal task. Top row: Target Gabor for the search task (left) and its power spectrum (right). Bottom three rows: the mean *t* statistic across four consecutive days of the difference between log Fourier power of fixated patches and shuffled controls, for monkeys MAS15 (left) and MAS 16 (right) for three different patch sizes; from top to bottom: 1° × 1°, 2° × 2° and 4° × 4°. Fourier power is clearly enhanced along the direction of target orientation. Units on the color bars indicate the mean *t* score.

at three different patch sizes (1° × 1°, 2° × 2° and 4° × 4°). We observed that the fixated patches have higher amplitude than nonfixated patches (positive *t* scores, Figure 5), suggesting that monkeys are biased to look at locations of high amplitude. As expected, high edge–energy patches are attractive targets for fixations.

Crucially, this same Fourier analysis allows us to inquire if the peaks in energy are consistent with the target orientation. Indeed, we also observed that the fixated patches have higher energy along the target orientation. In particular, we found higher energy along the vertical direction for the vertical task and along the horizontal direction for the horizontal task (Figure 5) and that this effect appears was more pronounced for 4° × 4° patches than 1° × 1° patches. The greater concentration of target power for the 4° × 4° case might be consistent with motor noise in executing saccades, where the saccade lands slightly away from the intended location. Thus, Fourier analysis provides a clear and intuitive visualization that monkeys fixate spots where the target orientation is dominant and this effect is better detected when visual features are averaged over larger areas in the periphery.

ROC Analysis

Fourier analysis suggested that energy and task-specific target orientation were diagnostic of fixated patches during natural scene search. Motivated by this observation, we intended to quantify the ability of visual features to predict fixations. We extracted local visual features: saliency, edge–energy, relevance, orientedness, and verticalness (see Methods) from fixated patches and shuffled controls, and then compared them using ROC analysis. We found that on average across animals and sessions for each task, fixated patches were more salient, had higher edge-energy, were more relevant, and had a higher verticalness than shuffled-control patches, regardless of the task; however, orientedness was not diagnostic of the difference between fixated and nonfixated patches[1] (Figure 6). According to the analysis of one feature at a time, saliency, edge–energy, relevance and verticalness are all predictive of fixations.

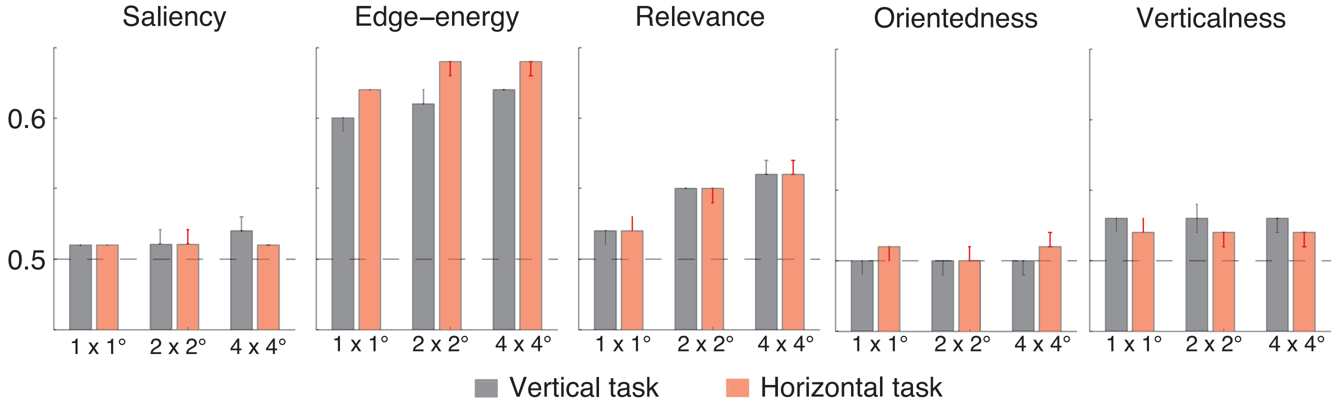

Figure 6. Area under receiver operating characteristic (ROC) curves (AUCs) comparing the distribution of visual features at fixated patches and shuffled controls across approximately 55,000 fixations (pooled across 16 sessions from two monkeys) for each task and each patch size. Black bars represent the vertical Gabor task and red curves represent the horizontal task. Error bars show bootstrapped 95% confidence intervals.

Evolution of search strategy within a trial

Our goal was to study the temporal evolution of search strategy at multiple time scales. To understand how saccade planning evolved at the relatively short time scale of within a single trial, we correlated the visual features at fixation with the fixation order (first, second, third, etc.) and the length of the saccade made to land at that fixation. We observed that fixated locations had different features at early and late fixations. In particular, saliency, edge-energy, relevance, orientedness, and verticalness of fixated patches increased from the beginning to the end of the trial (Figure 7 and Supplementary Figure S3; filled black diamonds; $p < 0.0001$ for both tasks). We also found that saliency, edge-energy, relevance and verticalness of fixated patches were negatively correlated, $p < 0.0001$ for both animals and tasks, with the length of the saccade made to get to that location (Figure 7 and Supplementary Figure S2; filled blue squares). By contrast, orientedness was positively correlated with fixation number and saccade length (Figure 7 and Supplementary Figure S2). These effects are abolished as expected when the same analysis is performed with shuffled controls instead of fixated patches (unfilled black diamonds and blue squares in Figure 7 and Supplementary Figure S2). Crucially, these effects are only consistent across animals and tasks when a retinal transform is applied (last data point in all panels of Figure 7 and Supplementary Figure S2). We further verified that these phenomena are independent; that is, although patches of high relevance or edge-energy are fixated early on in the trial by short saccades, this is not merely because early saccades are short on average. In particular, we found significant but very weak correlations between fixation order and saccade length for either monkey (across tasks: $r = -0.03$ for the vertical task and $r = 0.02$ for the horizontal task; $p < 0.01$ for both correlations), which were not sufficient to explain the high correlations observed with visual features at fixation. Together, these results show how search strategy evolves during the trial, with predictive visual features being fixated later in the trial.

Multivariate analysis of the evolution of search strategy over days

Although ROC analysis suggests that saliency, relevance and edge–energy can predict fixations, one fundamental limitation of the technique is that it can only quantify the effect of one variable at a time. However, in practice, these variables tend to be correlated. In our data, for instance, relevance is clearly correlated with edge–energy ($\rho = 0.25$; $p < 0.0001$; 16 sessions, two animals, 2° × 2° patch with retinal transform, 57,775 fixations) and edge–energy is correlated with saliency, $\rho = 0.02$; $p < 0.0001$. The correlation is problematic because the true effect of edge–energy might manifest as an apparent effect of saliency or relevance. To address this concern, we applied multivariate logistic regression to decode fixated from nonfixated patches as a function of local visual features (see Methods).

Visual features could discriminate fixated from nonfixated patches

Using multivariate logistic regression we asked if fixated patches could be decoded from shuffled-control patches. We found that fixated patches and shuffled controls could be decoded reliably above chance level (50%). Decoding accuracies did not differ significantly across tasks (vertical vs. horizontal target). Fixated patches were better decoded when peripheral visual

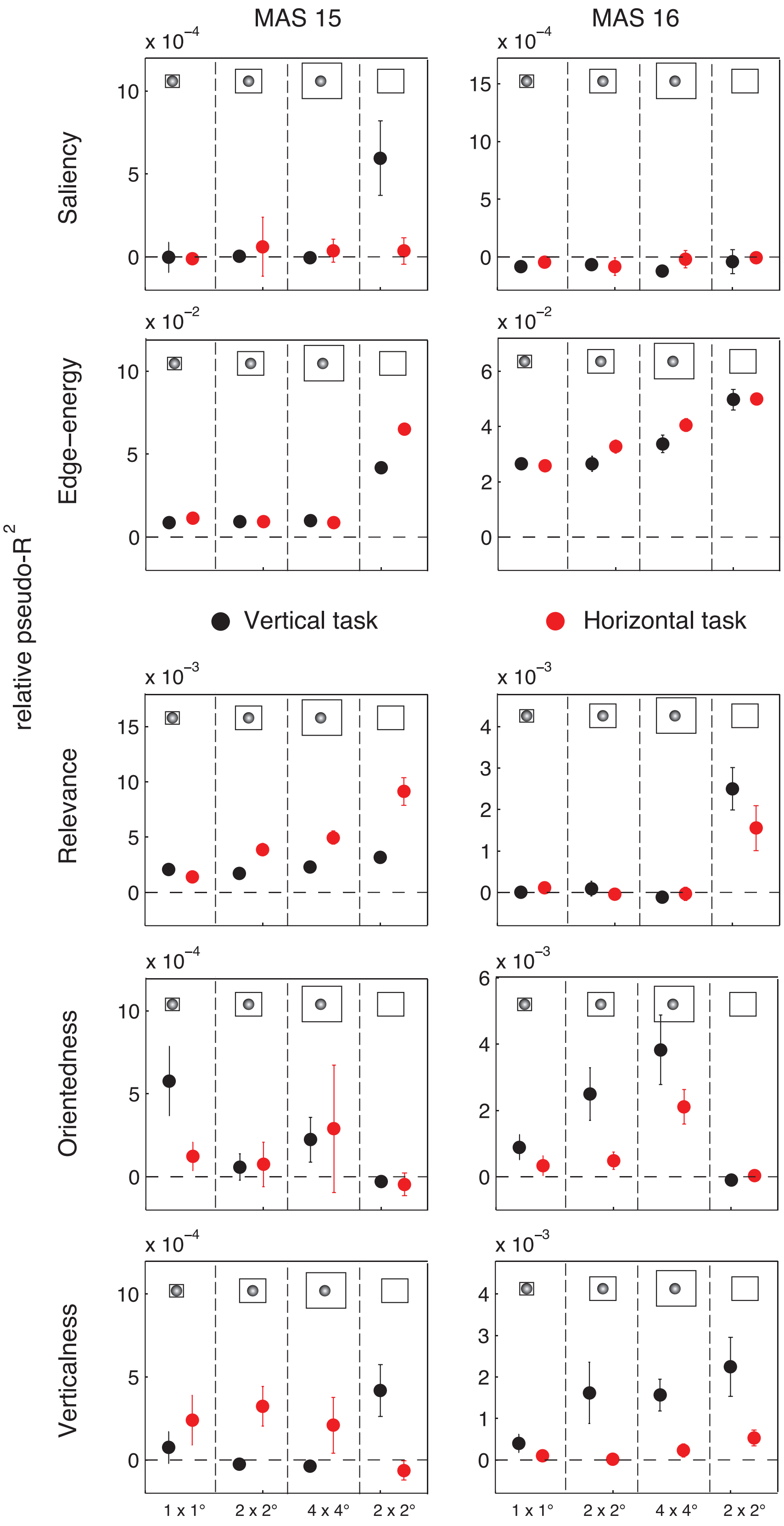

Figure 7. How does search strategy evolve within a trial? Spearman's rank correlation coefficients between fixation order (black diamonds), or saccade length (blue squares) and visual features: saliency, edge–energy, relevance, orientedness, and verticalness are shown for one monkey (MAS15) for each task. Filled shapes show effects for fixated patches and unfilled shapes for shuffled controls.

→

←

Error bars show bootstrapped 95% confidence intervals. The symbols on the *x*-axis indicate the image processing done prior to feature computation: the first three data points correspond to a retinal transform followed by averaging over 1° × 1°, 2° × 2°, and 4° × 4° windows respectively; the fourth data point indicates averaging over a 2° × 2° window without applying any retinal transform. The star indicates a significance level of $p < 0.0001$.

patches were not degraded by the retinal transform (last column of Table 1).

Relative influence of visual features on saccade choice

Having established that the visual features at fixation could predict eye movement behavior, we were interested in the relative influence of individual features. By decoding fixated patches from shuffled controls in the two tasks, we found that edge–energy was predominantly more influential than all other features in predicting fixations and saliency was completely explained away for both monkeys (top two panels in Figure 8; Table 2). These results appear to agree with findings in human studies which suggest that bottom-up saliency plays a minimal role in explaining saccades during search tasks (e.g., Einhäuser et al., 2008). However, it must be noted that if edge–energy is a bottom-up feature, then bottom-up features appear to play an important role in gaze prediction even during search.

Given the demands of the search task (finding a Gabor target in a natural scene), it is worth considering the possibility that edge–energy is not strictly a bottom-up feature, but that locations with high edge–energy are sought after by the monkeys simply because Gabor wavelets have high edge–energy. Although seeking high-energy patches is one possible search strategy, our Fourier analysis (Figure 5) and ROC analysis (Figure 6) suggest that target similarity (relevance) or target orientation also plays a role. Indeed, when using multivariate decoding, we found that task-relevant features had nonzero effects (bottom three panels in Figure 8; Table 2). In particular, relevance played an important role for MAS15 but not MAS16, whereas the reverse was true for orientedness (Figure 8, panels 3 and 4 from above; Table 2). On the basis of this result, it is not possible to conclude whether the internal search template guiding top-down saccade choice is comprised of the target in its entirety (relevance), orientedness, or the target orientation (verticalness); one dominant strategy does not win out across data from both animals and tasks, but all seem to be important.

Evolution of search strategy over multiple weeks

Although we did not find a change in search strategy over consecutive days, monkey MAS15 had performed the vertical search task ahead of the horizontal task, and the order of tasks was swapped for MAS16. Since these tasks were separated by a few weeks, we had the opportunity to compare search strategy across tasks. Indeed, for the 2° × 2° window, we observed a significant difference in the effect size of verticalness (a feature that captures the predominant orientation) across the two tasks. In particular, for MAS15, who had performed the horizontal task later, verticalness is significantly predictive of saccades in the horizontal task, with shuffled-control patches being more vertical that fixated patches ($\beta \pm SE = -0.0546 \pm 0.0161$; $p = 0.0009$; see Table 2B and bottom panels in Figure 8). For MAS16, who had performed the vertical task later, verticalness is significantly predictive of saccades in the vertical task, with fixated patches being more vertical than shuffled-control patches ($\beta \pm \mathrm{SE} = 0.1232 \pm 0.0201$; $p < 0.0001$; see Table 2C and bottom panels in Figure 8). This practice effect of verticalness also correlates with search performance as quantified by number of fixations required to find the target. In particular, a Wilcoxon rank-sum test for different median number of fixations for earlier and later tasks suggests that both monkeys improve across tasks ($p < 0.00001$; $z = 29.1$; median of 6 vs. 4 fixations for MAS15, and $p < 0.00001$; $z = 8.8$; median of 8 vs. 7 fixations for MAS16). From these results, it appears that animals gradually learn to place more emphasis on

	Retinal transform 1° × 1°	Retinal transform 2° × 2°	Retinal transform 4° × 4°	No retinal transform 2° × 2°
MAS15, vertical task	55.7 ± 0.9	55.5 ± 0.6	56.8 ± 0.5	61.5 ± 0.7
MAS15, horizontal task	56.5 ± 0.6	56.4 ± 0.8	56.6 ± 0.8	64.3 ± 1.0
MAS16, vertical task	58.4 ± 0.6	60.2 ± 1.3	60.5 ± 1.4	62.9 ± 0.7
MAS16, horizontal task	59.2 ± 0.8	60.9 ± 0.6	62.3 ± 1.3	62.3 ± 1.0

Table 1. Decoding accuracies (mean percentage of correctly predicted patches and standard deviation across 10 cross-validation folds) for fixated patches versus shuffled controls using all bottom-up and task-relevant visual features as inputs to a logistic regression.

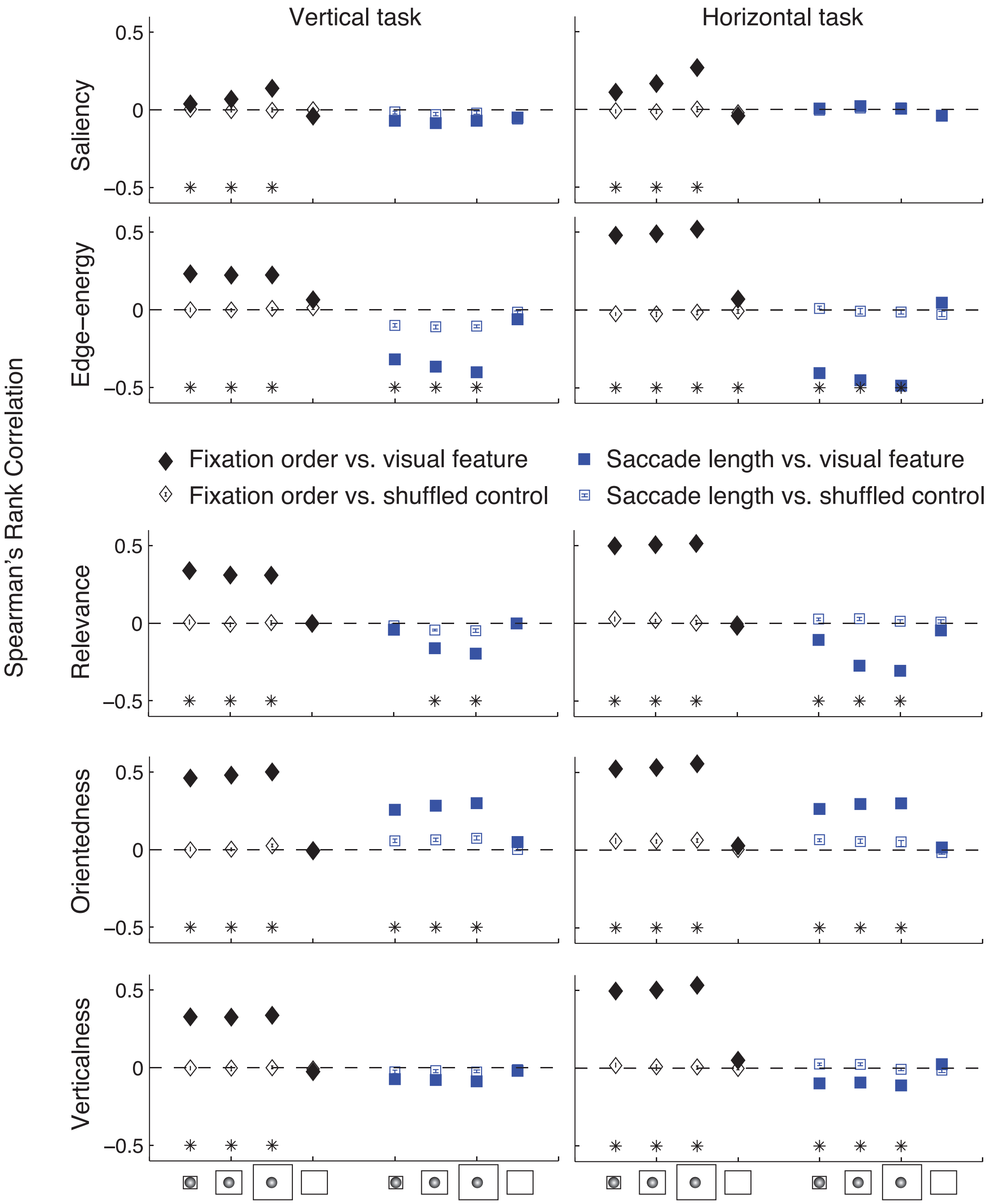

Figure 8. Relative influence of visual features on gaze behavior during search. The influence of each feature on eye gaze was obtained by measuring the extent to which a full model comprising all features would improve upon a partial model excluding that feature. Effect sizes (relative pseudo-R^2 on the test set) of features for the vertical (black) and horizontal (red) Gabor search tasks for each monkey. Error bars show standard errors of mean across 10 cross-validation folds.

the target orientation over a period of several weeks of performing the two related tasks. Furthermore, this long time-scale practice effect is not consistently detected by our analysis if we do not apply a retinal transform (bottom panels of Figure 8), suggesting once again that a biologically realistic model of peripheral vision is essential to understand eye movement behavior during search.

Discussion

We set out to examine the contents of the visual priority map employed in saccade guidance and how they evolved over time when monkeys searched for Gabor targets embedded in grayscale natural scenes. We applied multivariate analysis technique to tease out the relative contributions of various task-relevant

Variable	*β**	*SE*	*t*	*p*	Likelihood ratio	*p*
Constant	0.0180	0.0108	1.6706	0.0954		
Saliency	0.0085	0.0108	0.7809	0.4503	0.6797	0.9442
Edge-energy	0.2612	0.0131	19.8995	<**0.0001**	455.2963	**0**
Relevance	0.1154	0.0133	8.6978	<**0.0001**	83.1493	<**0.0001**
Orientedness	−0.0268	0.0120	−2.2281	0.0319	5.0343	0.3035
Verticalness	−0.0096	0.0126	−0.7664	0.4607	0.6666	0.9450

Table 2. Regression tables for logistic regression decoding of 2° × 2° fixated patches and shuffled controls after the retinal transform was applied. Separate regression tables for each animal (MAS15, MAS16) and task (vertical or horizontal Gabor search) provide coefficient estimates, standard errors, *t* scores and log-likelihoods, chi-squared (χ^2) statistics, and pseudo-R^2s ($R_D{}^2$), as well as likelihood ratios for all leave-one-feature-out partial models. Significant features are emphasized in boldface.

Part A. MAS15: Vertical target. LR $\chi^2(6) = 704.4$; $p = 0$; log-likelihood = −23,992; $R_D{}^{2**} = 0.0143$. * Standardized. ** Pseudo-R^2 (test set).

Variable	*β**	*SE*	*t*	*p*	Likelihood ratio	*p*
Constant	0.0337	0.0130	2.5918	0.0097		
Saliency	0.0287	0.0131	2.1859	0.0341	4.8705	0.3187
Edge-energy	0.2880	0.0177	16.2238	<**0.0001**	316.5875	**0**
Relevance	0.1889	0.0176	10.7034	<**0.0001**	131.5860	**0**
Orientedness	−0.0359	0.0163	−2.2012	0.0417	4.9799	0.3190
Verticalness	−0.0546	0.0161	−3.3997	**0.0009**	11.6278	**0.0258**

Part B. MAS15: Horizontal target. LR $\chi^2(6) = 704.4$; $p = 0$; log-likelihood = −16,714; $R_D{}^{2**} = 0.0203$. * Standardized. ** Pseudo-R^2 (test set).

Variable	*β**	*SE*	*t*	*p*	Likelihood ratio	*p*
Constant	0.0861	0.0165	5.2211	<0.0001		
Saliency	−0.0032	0.0164	−0.1958	0.7475	0.1470	0.9953
Edge-energy	0.5386	0.0253	21.2555	<**0.0001**	584.6116	**0**
Relevance	0.0380	0.0197	1.9298	0.0621	3.9210	0.4398
Orientedness	−0.1494	0.0198	−7.5346	<**0.0001**	57.1285	<**0.0001**
Verticalness	0.1232	0.0201	6.1210	<**0.0001**	37.5927	<**0.0001**

Part C. MAS16: Vertical target. LR $\chi^2(6) = 769.5$; $p = 0$; log-likelihood = −10,576; $R_D{}^{2**} = 0.0346$. * Standardized. ** Pseudo-R^2 (test set).

Variable	*β**	*SE*	*t*	*p*	Likelihood ratio	*p*
Constant	0.0775	0.0144	5.3812	<0.0001		
Saliency	−0.0069	0.0143	−0.4846	0.6132	0.4056	0.9694
Edge-energy	0.6241	0.0236	26.4207	<**0.0001**	953.3296	**0**
Relevance	0.0025	0.0164	0.1493	0.7738	0.1008	0.9982
Orientedness	−0.0707	0.0178	−3.9771	**0.0002**	15.9725	**0.0059**
Verticalness	0.0224	0.0173	1.2956	0.2320	1.8565	0.7618

Part D. MAS16: Horizontal target. LR $\chi^2(6) = 1{,}099.4$; $p = 0$; log-likelihood = −14,001; $R_D{}^{2**} = 0.0372$. * Standardized. ** Pseudo-R^2 (test set).

features while allowing for bottom-up features to explain them away. We found that over short time scales of a single trial, monkeys preferred to fixate target locations with high saliency, edge–energy, relevance and verticalness with short saccades and locations with high orientedness with long saccades. As the trial progressed, locations of higher bottom-up features, as well those of higher relevance for the search task, were more likely to be fixated. Over longer time scales of weeks, animals were able to adapt their search strategy to seek our task-relevant features at saccade targets, and this improvement in strategy was associated with statistically significantly fewer fixations required to find the target in successful trials. Crucially, our findings only reveal themselves once realistic models of peripheral vision are taken into account.

The multivariate decoding accuracies leave some anomalies to be explained. In particular, they suggest that patches fixated by MAS16, the worse-performing monkey are better discriminated than those fixated by MAS15. One potential explanation is that MAS16 relies on edge–energy more than MAS15 (Figure 8), which, as we have seen with the ROC analysis, (Figure 6) is the strongest predictor of fixations. Further, MAS16 show improved decoding with increased spatial extent of averaging. This might be because of the 2-D auto-correlation function of visual features around fixations; that of edge–energy might be more spread than that of saliency or relevance or other features.

One limitation of our approach is that although we have used multivariate regression to explain away correlated features, our data do not establish a causal relationship between certain visual features and gaze. For instance in our current design, since Gabor targets have high edge–energy and we have not parametrically varied edge–energy content across experimental sessions, it is not possible to say whether edge–energy is a bottom-up or a top-down factor. By parametrically manipulating individual features in the scene and comparing gaze behavior across these conditions, it is possible to establish a more direct relationship between visual features at fixation and factors that influence saccade choice (see e.g., Einhäuser et al., 2008; Engmann et al., 2009; ‘t Hart, Schmidt, Roth, & Einhäuser, 2013). In future studies, we intend to design experiments of this nature.

The natural world is seldom static and motion cues are known to be predictive of gaze (Le Meur, Le Callet, & Barba, 2007). Therefore, it might be argued that searching in static natural scenes may not be representative of natural eye movement behavior. However, it was our intention to design tasks that are as comparable as possible to studies in humans, so that our findings on monkeys would be directly comparable.

For the stimuli in this study, we have not taken into account the effect of generic task-independent biases that might inform eye movement behavior. For instance, it has been shown that human subjects tend to demonstrate a world-centered bias by often looking slightly below the horizon (Cristino & Baddeley, 2009). Future models would benefit from accounting for such ecological factors that bias gaze allocation.

An implicit assumption of our analysis framework—studying visual features at fixation to infer saccade-planning strategy—is that saccades are always made to planned locations that maximally resemble the target; that is, we assume that the oculomotor plan is a greedy strategy to maximize accuracy. This assumption has a couple of weaknesses. First, discrepancies between decision and action have been noted: it is known that saccades are sometimes made to intermediate locations between intended targets followed by subsequent corrective saccades (Findlay, 1982; Zelinsky, 2008; Zelinsky, Rao, Hayhoe, & Ballard, 1997). Second, it has been proposed that saccades might be made to maximize information (minimize Shannon entropy) rather than accuracy (Najemnik & Geisler, 2005); further, both Bayes-optimal and heuristic search models that maximize information seem to reproduce statistics of fixations made by human searchers better than a Bayesian model that only maximizes accuracy (Najemnik & Geisler, 2008, 2009). These are important phenomena that we have not addressed in our models of visual search.

Analysis of visual statistics at fixation

Although a large majority of priority map models (see models reviewed in Borji et al., 2013) have applied heuristics based on the knowledge of the visual system, a number of studies have analyzed visual information at fixated locations with the goal of reverse-engineering the features that attract eye gaze. These studies were done in humans but generally agree with our nonhuman primate’s results. We discuss a few pertinent human studies below in relation to our findings.

We found that edge–energy predicts fixations well. Given that principal components of natural images resemble derivatives of 2-D Gaussians (Rajashekar, Cormack, & Bovik, 2003), which are not very different from edge detectors, gazing at high-energy patches might be interpreted as maximizing the variance efficiently.

We found that relevance (resemblance to the target) could weakly but significantly predict fixations during search. These findings are in agreement with an earlier study showing that, when humans searched for simple geometric targets embedded in $1/f^2$ noise, the content of fixated locations resembled targets (Rajashekar, Bovik, & Cormack, 2006). They also showed considerable intersubject variation in search strategy, with two subjects selecting saccade targets based on the search-target shape, and one subject simply using a size criterion. We observe different search strategies in monkeys as well, with saliency and relevance (target similarity) predicting fixations to different relative extents in the two monkeys.

We observed clear difference in power spectra between fixated versus non-fixated patches, and also between these spectral differences across the two search tasks. However, it is important to note that power spectra predominantly capture second order statistics, and might therefore not capture the full extent of potential variance in fixated patches. Indeed, Krieger, Rentschler, Hauske, Schill, and Zetzsche (2000) showed that locations with larger spatial variance are more likely to be fixated, but this variance largely emerges from higher-order statistics quantifiable using bispectral densities and is therefore invisible in the power spectra. Their results provide a cautionary note to the interpretation of power-spectral differences that we observed in our analysis.

A number of studies in humans support our finding that the influence of bottom-up features is diminished while that of task-relevant features is enhanced during search in natural scenes (Castelhano & Heaven, 2010; Ehinger et al., 2009; Einhäuser et al., 2008; Henderson et al., 2007; Malcolm & Henderson, 2009; Schütz, Trommershäuser, & Gegenfurtner, 2012). Although, none of these studies have explicitly modeled the degraded representation of visual information in the periphery, they are in agreement with our finding that saliency gets explained away by edge–energy.

Studies that try to characterize the features humans use while searching or free-viewing artificial or natural stimuli are generally consistent with the results we obtain in our nonhuman primate study. In the light of these studies on humans, our study, while by no means methodologically novel, supports the existence of very similar eye-movement strategies in nonhuman primates. This is a necessary first step towards understanding the neural circuits that compute these decisions using invasive electrophysiology. An additional comparative study repeating the same task in humans might be a useful bridge between existing studies of natural scene search in humans and future results from primate electrophysiology studies.

The contents of the bottom-up priority map

Given that our experiment was a search task, and the fact that the influence of bottom-up features is diminished during search, our study might not be ideal to examine the contents of the bottom-up priority map. However, the explaining-away analysis built into multivariate logistic regression enabled us to tease apart the relative influences of a multiscale saliency metric and edge–energy.

Multivariate analysis might help resolve conflicting findings in many human studies of eye movement behavior (Baddeley & Tatler, 2006; Einhäuser & König, 2003; Reinagel & Zador, 1999). Although some of these studies involve free-viewing tasks and are therefore not directly comparable, our findings agree best with the study by Baddeley and Tatler (2006), who found that edge-detectors at two different spatial scales better predicted fixation behavior than contrast or luminance. In contrast, a free-viewing study comparing primate and human fixations (Berg et al., 2009) suggested that saliency triumphs over luminance contrast. Our approach to model peripheral degradation for feature extraction and to deploy an explaining-away technique for data analysis could potentially resolve this disagreement.

The contents of the top-down search template

A long-standing hypothesis in the area of visual search (Treisman & Gelade, 1980; Wolfe, 1994) suggests that search is guided by the matching of incoming visual input to an internal search template. The extent to which the internal template approximates the target is a matter of current debate. On the one hand, more information in the internal template seems to result in better search performance (e.g., Hwang et al., 2009). On the other hand, if the internal template is approximate, it allows for variability in target features across trials (e.g., Bravo & Farid, 2009). Although Bravo and Farid (2009) assessed various task-relevant candidate features (including similarity of luminance contrast, spatial frequency, orientation, and gradient of the search target) for their predictive power of fixation locations, their use of Pearson correlation coefficient and ROC areas prevents a fair comparison of the relative predictive power of correlated features. Further, the extent to which the search target can be described by low-level visual features also influences the internal template (Bravo & Farid, 2012; Reeder & Peelen, 2013) making generalizations across tasks difficult. In our study, the search target is a fixed Gabor patch that can be described easily by low-level features. Our results show that relevance (a metric that maximizes an exact match between the Gabor and the local visual features), orientedness, and the orientation-sensitive metric of verticalness can predict fixations to varying degrees in the two monkeys. Thus, even in the restricted context of our Gabor search task, it seems that monkeys adopt different internal templates comprising either all of the target features (relevance), or only a subset of its features (verticalness, horizontalness).

Conclusions

We studied the problem of why we look where we do by studying search in naturalistic conditions. By modeling peripheral visual degradation in retinocentric coordinates, and by using nuanced analysis methods, we were able to quantify the relative extents to which various bottom-up and task-relevant image features influenced eye movements. These analysis methods also

enabled us to distinguish between features that were important for fixations in general, and features that were specifically important for the search tasks studied. We showed how this search strategy evolves over time and how distinct factors are important at different time scales. Our study thus establishes essential groundwork to examine the neural representation of priority and the neural mechanisms by which a saccade decision is computed using neurophysiological methods in nonhuman primate models.

Keywords: nonhuman primates, eye movements, natural scenes, visual search, priority map, peripheral vision, saliency, edge–energy, relevance, orientation statistics

Acknowledgments

This work was funded by NIH Grant 1R01EY021579-01A1. We thank B. M. 't Hart for useful discussions and comments on the manuscript.

Commercial relationships: none.
Corresponding author: Pavan Ramkumar.
Email: pavan.ramkumar@northwestern.edu.
Address: Department of Physical Medicine and Rehabilitation, Northwestern University and Rehabilitation Institute of Chicago, Chicago, IL, USA.

Footnote

[1] For any feature, an AUC > 0.5 suggests that $M(\text{fixated}) > M(\text{controls})$.

References

Baddeley, R. J., & Tatler, B. W. (2006). High frequency edges (but not contrast) predict where we fixate: A Bayesian system identification analysis. *Vision Research, 46,* 2824–2833.

Berg, D. J., Boehnke, S. E., Marino, R. A., Munoz, D. P., & Itti, L. (2009). Free viewing of dynamic stimuli by humans and monkeys. *Journal of Vision, 9*(5):19, 1–15, http://www.journalofvision.org/content/9/5/19, doi:10.1167/9.5.19. [PubMed] [Article]

Bisley, J. W., & Goldberg, M. E. (2010). Attention, intention, and priority in the parietal lobe. *Annual Review of Neuroscience, 33,* 1–21.

Borji, A., Sihite, N. D., & Itti, L. (2013). Quantitative analysis of human–model agreement in visual saliency modeling: A comparative study. *IEEE Trans Image Processing, 23,* 55–69.

Bravo, M. J., & Farid, H. (2009). The specificity of the search template. *Journal of Vision, 9*(1):34, 1–9, http://www.journalofvision.org/content/9/1/34, doi:10.1167/9.1.34. [PubMed] [Article]

Bravo, M. J., & Farid, H. (2012). Task demands determine the specificity of the search template. *Attention Perception and Psychophysics, 74,* 124–131.

Castelhano, M. S., & Heaven, C. (2010). The relative contribution of scene context and target features to visual search in scenes. *Attention, Perception, & Psychophysics, 72,* 1283–1297.

Cristino, F., & Baddeley, R. J. (2009). The nature of visual representations involved in eye movements when walking down the street. *Visual Cognition, 17,* 880–903.

Ehinger, K. A., Hidalgo-Sotelo, B., & Torralba, A. (2009). Modeling search for people in 900 scenes: A combined source model of eye guidance. *Visual Cognition, 17,* 945–978.

Einhäuser, W., & König, P. (2003). Does luminance-contrast contribute to a saliency map for overt visual attention? *European Journal of Neuroscience, 17,* 1089–1097.

Einhäuser, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision, 8*(2):2, 1–19, http://www.journalofvision.org/content/8/2/2, doi:10.1167/8.2.2. [PubMed] [Article]

Engmann, S., 't Hart, B. M., Sieren, T., Onat, S., König, P., & Einhäuser, W. (2009). Saliency on a natural scene background: Effects of color and luminance contrast add linearly. *Attention, Perception, & Psychophysics, 71,* 1337–1352.

Fecteau, J. H., & Munoz, D. P. (2006). Salience, relevance, and firing: A priority map for target selection. *Trends in Cognitive Sciences, 10,* 382–390.

Fernandes, H. L., Stevenson, I. H., Phillips, A. N., Segraves, M. A., & Kording, K. P. (2014). Saliency and saccade encoding in the frontal eye field during natural scene search. *Cerebral Cortex, 24,* 3232–3245.

Findlay, J. M. (1982). Global visual processing for saccadic eye movements. *Vision Research, 22,* 1033–1045.

Ganguli, D., Freeman, J., Rajashekar, U., & Simoncelli, E. P. (2010). Orientation statistics at fixation. *Journal of Vision, 10*(7): 533, http://www.

journalofvision.org/content/10/7/533, doi:10.1167/10.7.533. [Abstract]

Geisler, W. S., & Perry, J. S. (1998). Real-time foveated multiresolution system for low-bandwidth video communication. In *Photonics West'98 Electronic Imaging*. 294–305. International Society for Optics and Photonics.

Gottlieb, J. (2012). Attention, learning, and the value of information. *Neuron, 76,* 281–295.

Gottlieb, J., Oudeyer, P. Y., Lopes, M., & Baranes, A. (2013). Information-seeking, curiosity, and attention: Computational and neural mechanisms. *Trends in Cognitive Sciences, 17,* 585–593.

Hays, A. V., Richmond, B. J., & Optican, L. M. (1982). Unix-based multiple-process system, for real-time data acquisition and control. *WESCON Conference Proceedings, 2,* 1–10.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., Mack, M. L. V. R., Fischer, M., Murray, W., & Hill, R.W. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R. P. G. van Gompel (Ed.), *Eye movements: A window on mind and brain* (pp. 537–562). Amsterdam, The Netherlands: Elsevier.

Hooge, I. T. C., & Erkelens, C. J. (1999). Peripheral vision and oculomotor control during visual search. *Vision Research, 39,* 1567–1575.

Hwang, A. D., Higgins, E. C., & Pomplun, M. (2009). A model of top-down attentional control during visual search in complex scenes. *Journal of Vision, 9*(5):25, 1–18, http://www.journalofvision.org/content/9/5/25, doi:10.1167/9.5.25. [PubMed] [Article]

Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience, 2,* 194–203.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20,* 1254–1259.

Kanan, C., Tong, M. H., Zhang, L., & Cottrell, G. W. (2009). SUN: Top-down saliency using natural statistics. *Visual Cognition, 17,* 979–1003.

Kano, F., & Tomonaga, M. (2009). How chimpanzees look at pictures: A comparative eye tracking study. *Proceedings of the Royal Society B, 276,* 1949–1955.

Kienzle, W., Franz, M. O., Schölkopf, B., & Wichmann, F. A. (2009). Center-surround patterns emerge as optimal predictors for human saccade targets. *Journal of Vision, 9*(5):7, 1–15, http://www.journalofvision.org/content/9/5/7, doi:10.1167/9.5.7. [PubMed] [Article]

Krieger, G., Rentschler, I., Hauske, G., Schill, K., & Zetzsche, C. (2000). Object and scene analysis by saccadic eye-movements: An investigation with higher-order statistics. *Spatial Vision, 13,* 201–214.

Kuo, R.-S., Kording, K., & Segraves, M. A. (2012). *Predicting rhesus monkey eye movements during natural image search*. New Orleans, LA: Society for Neuroscience.

Le Meur, O., Le Callet, P., & Barba, D. (2007). Predicting visual fixations on video based on low-level visual features. *Vision Research, 47,* 2483–2498.

Malcolm, G. L., & Henderson, J. M. (2009). The effects of template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision, 9*(11):8, 1–13, http://www.journalofvision.org/content/9/11/8, doi:10.1167/9.11.8. [PubMed] [Article]

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). New York: Academic Press.

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature, 434,* 387–391.

Najemnik, J., & Geisler, W. S. (2008). Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision, 8*(3):4, 1–14, http://www.journalofvision.org/content/8/3/4, doi:10.1167/8.3.4. [PubMed] [Article]

Najemnik, J., & Geisler, W. S. (2009). Simple summation rule for optimal fixation selection in visual search. *Vision Research, 49,* 1286–1294.

Navalpakkam, V., & Itti, L. (2006). An integrated model of top-down and bottom-up attention for optimal object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 2, (pp. 2049–2056).

Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. In *Journal of the Royal Statistical Society, Series A (General)*, (pp. 370–384).

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision, 42,* 145–175.

Phillips, A., & Segraves, M. (2010). Predictive activity in the macaque frontal eye field neurons during natural scene searching. *Journal of Neurophysiology, 103,* 1238–1252.

Rajashekar, U., Bovik, A. C., & Cormack, L. K. (2006). Visual search in noise: Revealing the influence of structural cues by gaze-contingent classification image analysis. *Journal of Vision, 6*(4): 7, 379–386, http://www.journalofvision.org/content/6/4/7, doi:10.1167/6.4.7. [PubMed] [Article]

Rajashekar, U., Cormack, L. K., & Bovik, A. C. (2003). Image features that draw fixations. In *Proceedings of the International Conference on Image Processing,* 2 (pp. 313–316).

Reeder, R. R., & Peenlen, M. V. (2013). The contents of the search template for category-level search in natural scenes. *Journal of Vision, 13*(3):13, 1–13, http://www.journalofvision.org/content/13/3/13, doi:10.1167/13.3.13. [PubMed] [Article]

Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network, 10,* 341–350.

Renninger, L. W., Verghese, P., & Coughlan, J. (2007). Where to look next? Eye movements reduce local uncertainty. *Journal of Vision, 7*(3):6, 1–17, http://www.journalofvision.org/content/7/3/6, doi:10.1167/7.3.6. [PubMed] [Article]

Schütz, A. C., Trommershäuser, J., & Gegenfurtner, K. R. (2012). Dynamic integration of information about salience and value for saccadic eye movements. *Proceedings of the National Academy of Sciences, 109*(19), 7547–7552.

Serences, J. T., & Yantis, S. (2006). Selective visual attention and perceptual coherence. *Trends in Cognitive Sciences, 10,* 38–45.

‘t Hart, B. M., Schmidt, H. C. E. F., Roth, C., & Einhäuser, W. (2013). Fixations on objects in natural scenes: Dissociating importance from salience. *Frontiers in Psychology, 4.*

Tatler, B. W., & Vincent, B. T. (2009). The prominence of behavioral biases in eye guidance. *Visual Cognition, 17,* 1029–1054.

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems, 14,* 391–412.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136.

Tseng, P. H., Carmi, R., Cameron, I. G. M., Munoz, D. P., & Itti, L. (2009). Quantifying center bias of observers in free viewing of dynamic natural scenes. *Journal of Vision, 9*(7):4, 1–17, http://www.journalofvision.org/content/9/7/4, doi:10.1167/9.7.4. [PubMed] [Article]

Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks, 19,* 1395–1407.

Wilming, N., Harst, S., Schmidt, N., & König, P. (2013). Saccadic momentum of facilitation of return saccades contribute to an optimal foraging strategy. *PLoS Computational Biology, 9,* e1002871.

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review, 1,* 202–238.

Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review, 115,* 787–835.

Zelinsky, G. J., Adeli, H., Peng, Y., & Samaras, D. (2013). Modeling eye movements in a categorical search task. In *Philosophical Transactions of the Royal Society B, 368,* 20130058.

Zelinsky, G. J., Rao, R., Hayhoe, M., & Ballard, D. (1997). Eye movements reveal the spatio-temporal dynamics of visual search. *Psychological Science, 8,* 448–453.

Appendix A: Logistic regression

Model

Logistic regression is ideal for two-class classification problems, with the two classes in our problem represented by fixated patches and shuffled controls, respectively.

In logistic regression, the output is modeled as a Bernoulli random variable with mean probability of successful outcome given by

$$p(y=1|X,\beta)=\frac{1}{1+e^{-X\beta}} \tag{A1}$$

A Bernoulli random variable represents the outcome of a biased coin toss: in our problem, the potential outcomes of the toss represent the possibilities of whether the patch was truly fixated or whether it was a random shuffled control.

$$y \sim Bernoulli\{p(y=1|X,\beta)\} \tag{A2}$$

The probability of the outcome is then defined by the joint influence of independent variables in the design matrix X. Here, each column of X comprises the visual features extracted from the image patches centered at each fixation, and each row is a fixation. A bias term is also included.

The regression problem is then solved by estimating the coefficients β that minimize the negative log likelihood of the model, given by

$$-L=-\sum_i y_i \log(p_i)-\sum_i(1-y_i)\log(1-p_i) \tag{A3}$$

Unlike linear regression where the residuals are normally distributed and a closed-form solution exists, logistic regression is a convex problem in that it is typically solved using gradient-based methods. In practice, we implemented the method using Matlab's glmfit function.

Goodness of fit and model comparison using likelihood ratios

We computed the likelihood ratios between the model under consideration and a null model with only a bias term. The likelihood ratios are defined in terms of the model deviances, given by:

$$D_0 = -2\log(\mathcal{L}_0/\mathcal{L}_{sat}) \tag{A3}$$

$$D = -2\log(\mathcal{L}/\mathcal{L}_{sat}) \tag{A4}$$

where $\mathcal{L}_0$, $\mathcal{L}_{sat}$, and $\mathcal{L}$ are the likelihoods of the null model (one that only predicts the bias in the data), the saturated model (one that perfectly predicts the data), and the model under consideration. The likelihood ratio is then given by:

$$D - D_0 = -2\log(\mathcal{L}/\mathcal{L}_0) \tag{A5}$$

A negative likelihood ratio suggests that the model under consideration significantly outperforms the null model. The likelihood ratio is distributed as a chi-squared statistic and can therefore be converted into a goodness of fit χ^2 measure, given the number of independent variables in the model.

Relative pseudo-R^2

A related metric to the likelihood ratio is the pseudo-R^2. The idea of the pseudo-R^2 metric is to map the likelihood ratio into a [0, 1] range, thus offering an intuition similar to the R^2 for normally distributed data.

Many definitions exist for the pseudo-R^2, but we used McFadden's formula: ${R_D}^2 = 1 - L/L_0$, where L is the log-likelihood of the model under consideration and L_0 is the log-likelihood of the baseline (intercept-only) model (McFadden et al., 1974).

To estimate the relative effect size of each feature, or a subset of features, we used a leave-one-feature-out (leave-some-features-out) technique, a variant of best subsets of stepwise regression techniques. For each feature, we fit a partial model comprising all but that particular feature (or subset) and computed the pseudo-R^2 of the partial model. Then, to obtain a measure of the relative increase in predictive power obtained by adding that feature (or subset) back to the partial model, we used a measure called relative pseudo-R^2 (Fernandes et al., 2013). The relative measure is defined as $1 - L_{\text{full}}/L_{\text{partial}}$, where L_{full} and L_{partial} are the log-likelihoods of the full and partial models. The confidence intervals on this statistic were obtained by computing standard errors on 10 cross-validation folds; the relative pseudo-R^2 in each case was computed on the held-out test set.