

Where to look next? Eye movements reduce local uncertainty

Laura Walker Renninger

The Smith–Kettlewell Eye Research Institute,
San Francisco, CA, USA



Preeti Verghese

The Smith–Kettlewell Eye Research Institute,
San Francisco, CA, USA



James Coughlan

The Smith–Kettlewell Eye Research Institute,
San Francisco, CA, USA



How do we decide where to look next? During natural, active vision, we move our eyes to gather task-relevant information from the visual scene. Information theory provides an elegant framework for investigating how visual stimulus information combines with prior knowledge and task goals to plan an eye movement. We measured eye movements as observers performed a shape-learning and -matching task, for which the task-relevant information was tightly controlled. Using computational models, we probe the underlying strategies used by observers when planning their next eye movement. One strategy is to move the eyes to locations that maximize the total information gained about the shape, which is equivalent to reducing global uncertainty. Observers' behavior may appear highly similar to this strategy, but a rigorous analysis of sequential fixation placement reveals that observers may instead be using a local rule: fixate only the *most* informative locations, that is, reduce local uncertainty.

Keywords: eye movements, information theory, uncertainty, saliency, Bayesian statistics, shape discrimination, computational model

Citation: 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://journalofvision.org/7/3/6/>, doi:10.1167/7.3.6.

Introduction

Vision is more than a passive sense that is processed through a bottom-up hierarchy. We move our eyes, often purposefully, to actively gather the sensory information we need to complete a task. Our internal goals influence our eye-movement behavior, attention, and, ultimately, what we perceive and remember. How do we decide where to look?

The final landing point of a saccade is biased by mechanical and physiological constraints, but more important, the properties of the stimulus play a strong role in observed fixation locations. For example, saccades to a simple shape or object often land near the centroid of that object (Melcher & Kowler, 1999; Vishwanath & Kowler, 2003). When viewing natural images, observers tend to fixate regions with higher local contrast, such as regions near object borders or edges (Reinagel & Zador, 1999).

Due to the role of stimulus properties, researchers have approached the question of eye-movement planning by asking “what properties in the visual image draw fixations?” A popular hypothesis is that we look at salient points. Saliency is defined by visual features that stand out or are surprising because they have different brightness,

color, orientation, or motion than the surrounding features (Itti & Baldi, 2005; Itti & Koch, 2000). This hypothesis makes sense in terms of survival (we want to quickly locate the ripe fruit in the tree, or the tiger that moves suddenly in our periphery), in passive viewing situations such as watching a television commercial, or, perhaps, when planning the first saccade to a complex scene or object. However, because it is purely stimulus defined, saliency has limited applicability to how we actively use eye movements in daily life to recognize certain objects by their shape, parts, or feature configuration, to search for something, or to reach for an item.

The active, task-dependent nature of observed fixation locations was cleverly demonstrated by Yarbus (1967), who had observers study a painting with different questions in mind. In recent years, advancements in eye-tracking technology have allowed us to study eye movements as observers perform real-world tasks like making a sandwich or fixing a cup of tea (Hayhoe & Ballard, 2005; Hayhoe, Shrivastava, Mruczek, & Pelz, 2003; Land, Mennie, & Rusted, 1999). In these studies, eye movements provide “just-in-time” information relevant to the motor task about to be executed. These studies lend support to our intuition that eye movements are made to collect task-relevant information from the visual scene. How might we formalize this idea?

Information-theoretic approaches

Information theory provides a convenient framework in which we may formalize our intuitions and observations of eye-movement behavior. For example, both behavioral and physiological studies suggest that once a saccade has been made to a location, there is an active inhibition of return (IOR) signal that prevents revisiting that location again for a period of time (Dorris, Klein, Everling, & Munoz, 2002; Posner & Cohen, 1984). Within an information framework, the rationale for IOR becomes obvious—the fovea is not likely to return to a location once the information there has already been gathered. During sandwich making (Hayhoe et al., 2003) and object search (Rao, Zelinsky, Hayhoe, & Ballard, 2002), observers may land at average positions between potential target objects before moving to the final desired target. This average positioning or hedging provides more information about potential targets before the final eye-movement decision is made.

The tendency to saccade to regions of high local contrast (Reinagel & Zador, 1999) may also have an information-theory explanation. As Raj, Geisler, Frazor, and Bovik (2005) demonstrated, taking samples (fixations) that minimize contrast entropy provides the best information for the image reconstruction of natural scenes. To date, no direct comparison of human eye movements and entropy minimization in natural scenes has been made.

Legge, Hooven, Klitz, Mansfield, and Tjan (2002) and Legge, Klitz, and Tjan (1997) were among the first to use information theory to predict human eye-movement patterns in their ideal observer model for reading. Because eye movements are a natural consequence of having a foveated visual system, this architecture must figure prominently in any eye-movement model. As a stand-in for the falloff in resolution with eccentricity, Legge et al. defined the perceptual span—the number of letters available for processing during the current fixation. This “window” was used to make predictions about what word was being viewed and to plan the next saccade to a location that would best disambiguate the current word from other words in a well-defined vocabulary. Rather than attempting to incorporate the vast vocabulary of an individual, they shrewdly defined a lexicon of 542 common words. The model produced familiar reading patterns such as skipping small words and backward saccades; however, a fixation-by-fixation analysis was not performed.

Some recent attempts have been made to directly compare information-theory predictions against human eye-movement data (Najemnik & Geisler, 2005; Renninger, Coughlan, Verghese, & Malik, 2005). For example, how do observers use information during visual search? Najemnik and Geisler (2005) designed a simple yet clever search experiment. First, they carefully measured the visibility of a Gabor target in $1/f$ noise at various eccentricities. Using these measurements, they imple-

mented a search model for the Gabor target in noise that adopts one of three strategies: (1) move to random locations, (2) move to locations that reduce global uncertainty about target location, and (3) move only to locations that are most likely to contain the target (reduce local uncertainty). The second strategy will collect information optimally, and the third is the maximum *a posteriori* (MAP) strategy. The probability of target presence is monitored at every location, and the target is “found” when probability at one location exceeds a predetermined threshold. The authors demonstrate that their optimal and MAP searchers locate the target with roughly the same number of fixations as human observers. Although the aggregate behavior of human fixations qualitatively resembled their model fixations, the landing of individual saccades during the task was not examined. We have yet to unravel what decision strategies underlie the choice of human fixation locations.

Our approach

In this article, we use information theory to probe the underlying decision strategies that govern eye-movement planning. We use a psychophysical experiment that controls the observer’s task and the task-relevant visual information, as we measure eye movements. *Individual fixations* are compared against strategy predictions using a signal detection theory approach. At first inspection, human eye movements appear “optimal” (reduce global uncertainty); however, our rigorous analysis of individual fixation placement reveals that an approximate, local rule may actually govern eye-movement decisions.

Methods

Psychophysical methods

Stimuli

Observers participated in a shape-learning and -matching task. The shapes were novel and abstract silhouettes, created by randomly rotating and superimposing four randomly selected objects from the Snodgrass and Vanderwart (1980) data set of common objects. The motivation of this approach was to give the shape a mostly “natural” object-like boundary while still presenting an unrecognizable shape. By using abstract shapes in isolation, we avoid the influence of cognitive information on eye-movement planning, such as object familiarity or scene context, while still capturing edges—a fundamental feature of natural images. For each shape, a partner shape was created by superimposing a fifth randomly selected and rotated object from the Snodgrass data set. This action may cause more than one new protrusion on the original shape; however, only one single connected change was

kept. Some hand selection of the stimuli was performed to discard featureless or circular shapes. Our stimuli and stimulus generation code are available on the first author's web site. Five hundred shape pairs were created for the experiment. Stimuli were presented using Matlab with PsychToolbox (Brainard, 1997) and were presented as high-contrast white figures on a midgray background.

Design

We block the shape pairs into five levels of difficulty depending on the degree of boundary change (Figure 1A). We calculate the shape-pair difference as the change in orientation entropy along the boundary using a “fixation” at the shape centroid (see Appendix). We have shown that this shape difference quantity scales with human shape discrimination performance (Renninger, Verghese, & Coughlan, 2005a). We are currently working on a more detailed analysis of this metric.

Trials were completed in blocks of 20. Within each block, trial difficulty was balanced based on the amount of change between shape pairs. By including a range from easy to difficult trials in each block, subjects were motivated to continue studying the shapes effectively. Each trial contained a completely new shape pair; thus, no shape was seen more than once.

Task

Observers were seated 80 cm from the display screen, which subtended a visual angle of 22° . On each trial, observers first fixated a marker on the left or right side of the screen and pressed a button to begin. For the learning phase, a shape was selected at random from the pair and was displayed on the opposite side, centered at 10° from fixation. Subjects were instructed to maintain fixation at the marker. After 300 ms, fixation was extinguished and observers were allowed to make eye movements to and around the shape, which was displayed for 1,200 ms. Each displayed shape was scaled to measure 12.5° along the diagonal of its bounding rectangle.

The subject's task is to “learn” the shape during this brief presentation, and its large size ensures that eye movements are required to do so. In the matching phase, the learned shape and its highly similar partner were displayed together at a new location. These test shapes subtended 4° . Observers were allowed to study the shape pair until they reached a decision about which member of the pair matched the shape they learned (Figure 1B). Feedback was given.

Subjects

Four female subjects (of whom, three were naïve) participated in the experiment. Subjects ranged in age from 21 to 43 years. All had normal, uncorrected vision. A total

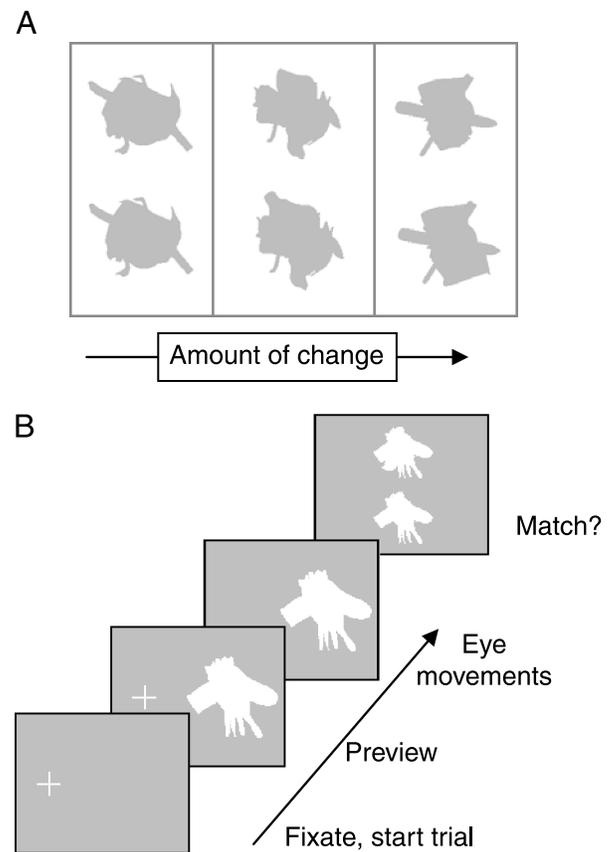


Figure 1. Stimuli and task. (A) Examples of shape pairs used in the psychophysical task. The size of the change increases from left to right. (B) Subjects fixate a marker and maintain fixation as a novel object silhouette appears in the periphery for 300 ms. When the marker is extinguished, the subject has 1,200 ms to study the shape with eye movements. Immediately after the study phase, the shape pair is displayed and the subject must select which one was just presented.

of 500 trials were completed over two to three recording sessions. The experimental protocol was approved by the California Pacific Medical Center Institutional Review Board.

Eye-tracking methods

Observers' eye movements were monitored during the task with an SRI Dual Image Purkinje Eye Tracker, sampling eye position at 1000 Hz. Viewing was binocular, but only the right eye was tracked. Head position was maintained with a bite bar. Calibration under these conditions is very stable, with high precision and no drift.

The calibration procedure was twofold. First, observers viewed a static cross that was made of 0.1° dots placed centrally, $\pm 5^\circ$ vertically, and ± 5 and $\pm 10^\circ$ horizontally.

Observers were instructed to view each dot in turn as the tracker was manually adjusted to return a linear readout in response to eye position. Next, observers fixated a dynamic 0.25° dot that blinked on for 1,500 ms and swept out a 5×5 calibration grid that covered the stimulus space in steps of 5° horizontally and 3.5° vertically. This grid was used to apply a piecewise perspective transformation to the raw eye position data, correcting any nonlinearity that may occur near the edges of the display.

The initiation of a saccade was marked if eye velocity exceeded $80^\circ/\text{s}$. The initiation of a fixation was marked if eye velocity dropped below $10^\circ/\text{s}$. Eye velocity was computed as the rate of displacement within a symmetric 10-ms window centered on the sample of interest. If two sequential fixations were found within 0.5° , they were labeled as a single fixation.

Modeling information

The observers' implicit task is to build an accurate representation of each shape as it is studied with eye movements so that it can be discriminated from a highly similar shape during the matching phase. Given our knowledge of V1 processing, we assume that the information needed for this task is the edge orientations derived from the shape contour. It is possible that a pixel-based representation would work as well given our task, but it is more computationally intensive. "Higher level" representations such as parts do not appear to play a role for this task (Renninger, Verghese, & Coughlan, 2006).

We constructed a straightforward model to describe how orientation information is gathered by the human visual system. With each fixation, the observer takes a foveated measurement of the orientations in the stimulus. The ability to resolve orientations degrades as a function of eccentricity. We estimate the orientation information at a point on the stimulus by constructing a pooling neighborhood whose size depends on distance from the current fixation point (Figure 2B). The pooling extent was determined using parameters from the vernier acuity literature (Levi, Klein, & Aitsebaomo, 1985; see Appendix) and has been confirmed by the authors in a separate, unpublished study. We chose these parameters because vernier acuity is thought to be a consequence of orientation-selective filters in primary visual cortex. For simplicity, we consider filters selective to eight discrete orientations. Within a pooling neighborhood, we count the number of occurrences of different veridical edge orientations and create a histogram (or probability distribution after normalization) of the different orientations at that location (Figure 2C). Each histogram is intended to be analogous to the initial distribution of neural responses to the stimulus across a hypercolumn of orientation-selective cells in visual cortex (Lee & Yu, 2000).

With each successive fixation, we update what is known about the stimulus at each point (posterior distribution) by

multiplying the new measurement distribution (likelihood) at that point with the prior distribution, which is flat before the very first fixation is made (Figure 2A). An adjustment of response distributions such as this may be achieved in the visual system through feedback connections (Hamker, 2003). By definition, information is the entropy of a probability distribution:

$$\text{entropy} = -\sum p(x)\log p(x). \quad (1)$$

When there are many different orientations in a neighborhood (e.g., a bumpy contour in the periphery), all orientations are equally likely and the distribution will be flat (high entropy). Alternatively, straight edges will produce energy at a single orientation or very peaked distributions (low entropy). As the evidence of orientations accumulates with successive fixations, we can represent the uncertainty of shape knowledge at any point in time by computing a resolution-dependent entropy (RDE) map (Figure 2D). We have shown in previous work that the residual uncertainty (entropy) remaining after a series of fixations correlates with observer performance in this task, validating our modeling approach (Renninger, Verghese, & Coughlan, 2005b). A more rigorous treatment of the model can be found in the Appendix.

In the next section, we first look at the general pattern of human eye movements in our task before applying the model to probe the nature of decision strategies that underlie individual fixations.

Experimental results

Task performance and eye movements

All four observers performed above chance but not perfectly in the shape-matching phase of the experiment. Percentage correct ranged from 75% to 78%. This performance level suggests that the task was achievable yet difficult enough to encourage efficient information gathering during the learning phase.

Mean amplitudes of object-exploring saccades ranged from 2.38° to 4.44° and rarely exceeded 10° . Mean dwell times ranged from 175 to 403 ms (Figures 3A and 3B). These measured eye-movement behaviors are similar to what has been found for viewing of naturalistic stimuli (Bahill, Adler, & Stark, 1975) and in search tasks (Najemnik & Geisler, 2005). There was a strong negative correlation between average saccade amplitudes and dwell time across observers ($r = -.90$). That is, observers who made large saccades tended to fixate for shorter periods and vice versa. Subjects typically made three to five fixations around the object in the viewing time allowed.

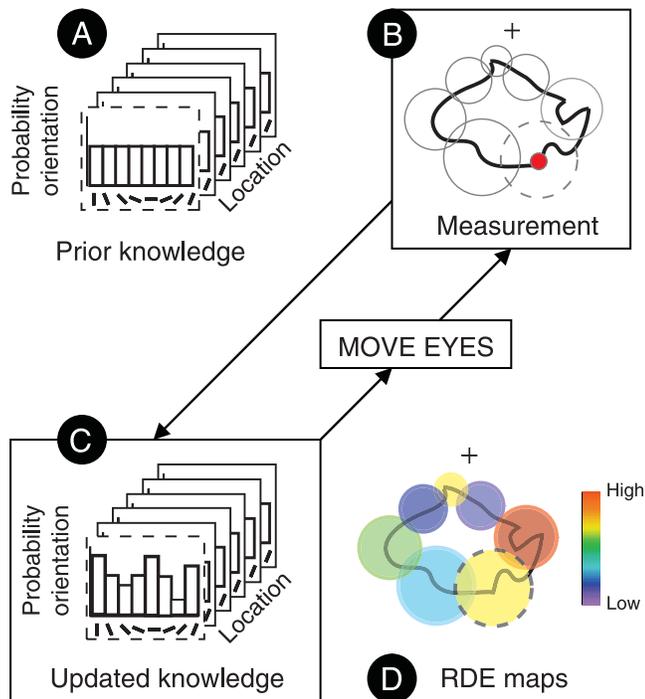


Figure 2. Probabilistic model of shape contour information. (A) Before the shape is investigated with eye movements, there is no prior knowledge about the orientation at each location in the stimulus space. The resulting probability distributions over orientation are flat at each location, and uncertainty (entropy) is high everywhere. (B) A sample fixation (+) places smaller pooling neighborhoods near the top of the shape and larger neighborhoods near the bottom. Orientation distributions are computed at a location (red dot), using the appropriate pooling area (dashed circle). (C) The measurement distribution is multiplied with the prior distribution at that location to produce updated knowledge (posterior distribution). Updated knowledge becomes prior knowledge for the next fixation. The eyes move and another measurement is taken. (D) Schematic. The uncertainty (or information) at any point in space and time is computed from the updated knowledge and can be represented with an RDE map. For the first fixation, straight lines within a pooling neighborhood result in lower entropy (blue) at a location, whereas curved or bumpy lines within a neighborhood result in higher entropy (red). See Appendix.

Fixated locations were found to be spatially distributed in a “donut” shape for three of four subjects (Figure 3C). These distributions were assembled by assuming that the preview fixation was always to the right and by flipping fixation coordinates for left-preview trials. The fixations that are highlighted in red are the *first* fixations to the object. First fixations do not have the same donut distribution of the other object-exploring fixations and are biased in the preview direction. This clustering of the first fixation for very different shapes may indicate that it may simply be a localizing saccade that is mostly

independent of detailed shape information. The absolute scale of the donut distribution might suggest that observers are making fixations within object boundaries. Further analysis revealed that although fixations may cluster near boundaries on average, they often fall outside of the boundary; 8.4% to 27.6% of fixations landed outside object boundaries depending on observer. Next, we evaluate different eye-movement decision strategies by examining the placement of *individual fixation locations*.

Strategy analysis

Do observers maximize information?

Using the information-theoretic model, we can probe how information is used to plan eye movements to the stimulus. If the goal of eye movements is to gather task-relevant information, then the best strategy is obvious: fixate locations that maximize the total information gained about the contour orientations. We can compute this prediction by evaluating all possible next fixation locations and selecting the one that yields the greatest gain in total

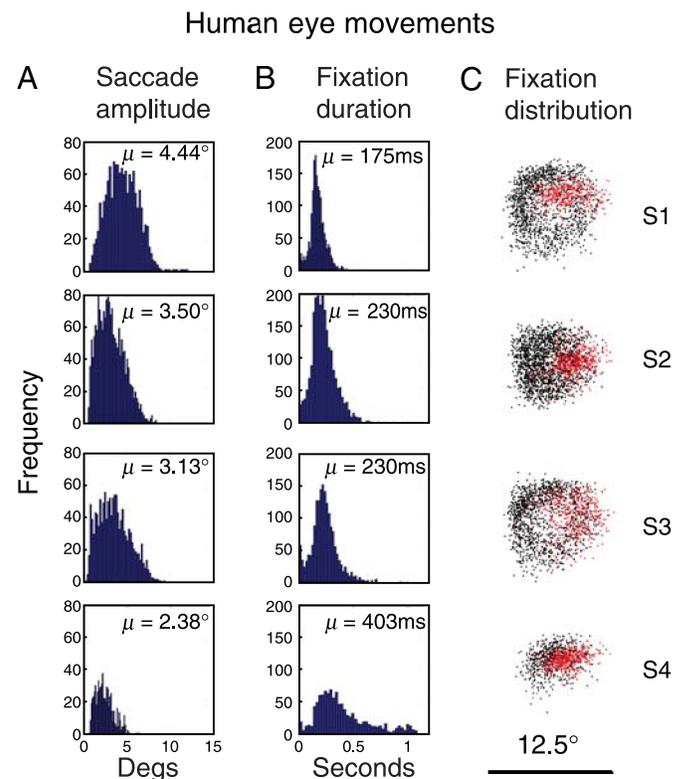


Figure 3. Human eye movements. (A) Human saccade amplitudes and (B) fixation durations are similar to what has been found in previous work. (C) Fixations tend toward the edges of the shapes for three of four observers, forming a donut-shaped distribution. Red points indicate first fixations to the shape.

information (i.e., greatest reduction in total uncertainty). We consider a grid of possible fixations, spaced 0.25° apart, and compute a strategy prediction map. **Figure 4A** illustrates a sequence of fixations based on this “global” strategy prediction. Note that this prediction is for placing the *next* fixation. Predicting the fixation *sequence* that maximized information gain is more computationally intensive, although there is some evidence that humans may indeed plan more than one fixation at a time.

When we consider the aggregate behavior of fixation predictions generated by the global strategy, we see that the distributions of saccade amplitude and fixation location are qualitatively similar to those measured for our subjects (compare **Figure 4B** with **Figure 3**). In contrast, the distributions generated by a random strategy that predicts fixations anywhere on the stimulus are quite different from the human pattern (**Figure 4C**).

Data exclusion

Recall that, on each trial, the observer first fixates the marker, followed by a second fixation on or near the shape. The few trials in which this pattern is violated are not included in the analysis. For the remaining trials, we wish to analyze only “object-exploring” fixations, which fall within the donut-shaped distribution and may be predicted by our information strategy. We do so by excluding the preview and localizing fixations, which will not be predicted by

information strategies. We also exclude fixations with dwell times that are less than 50 ms, as they do not fall within the primary mode of the population distribution (see **Figure 3B**) and may be stutters or pauses along the way to the intended fixation location.

Fixation error

To quantitatively assess each strategy, we first compute the “fixation error” as defined here. We compute the first five strategy-predicted locations for each shape. We chose the first five because human observers typically made three to five fixations per shape. Every human fixation to that same shape is then mapped to the closest strategy fixation, and the distance errors are accumulated. The mean of these samples is the fixation error and is taken as one measure of how well strategy-predicted locations align with human fixations. The significance of the alignment is assessed by bootstrapping (1,000 iterations) to get 95% confidence intervals of the fixation error. On average, human fixations are closer to the global strategy than to random fixations (**Figure 5**).

Note that this error measure ignores the sequence in which fixations are made. Nonetheless, it is useful for assessing the extent to which observers and strategies select the same “interest points” regardless of their strength (and, thus, order in the sequence). It also affords us a rough measure of how similar the fixation distributions are

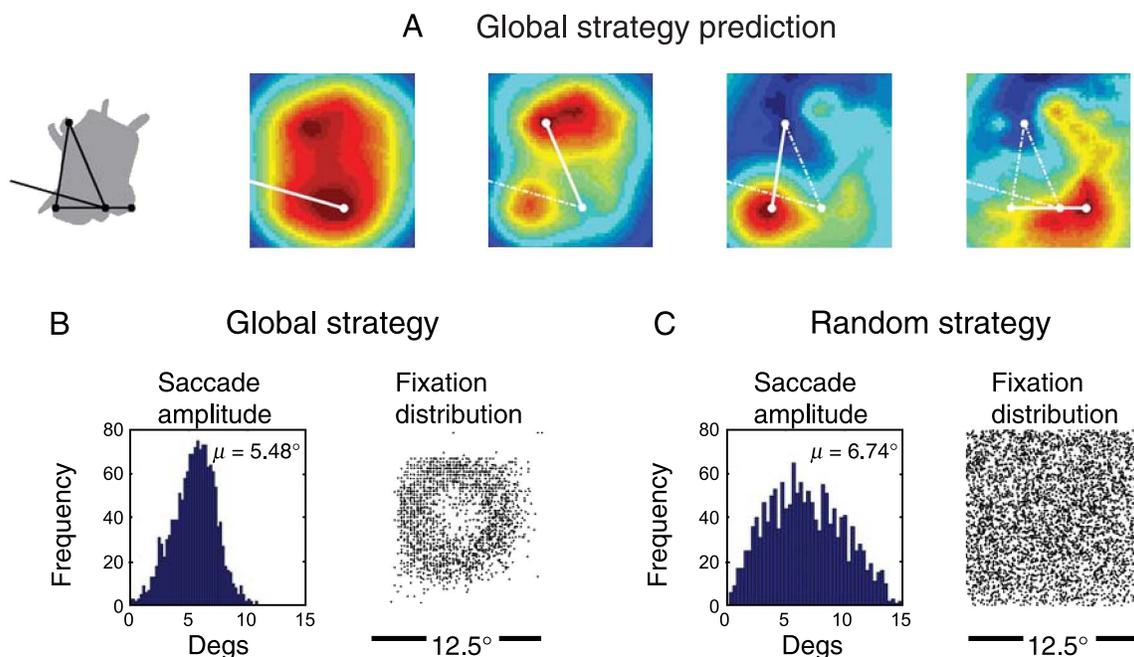


Figure 4. Global uncertainty and random strategy predictions. (A) Predicted fixations for the global strategy are superimposed on a shape stimulus (left) and the strategy prediction map (scaled from 0 to 1; blue to red). After each fixation, knowledge of the stimulus is updated and a new prediction is computed. The global strategy is to move to locations that maximize the total information gained (i.e., reduce uncertainty about all edge orientations) with each fixation. (B) The global strategy predicts saccade amplitude and fixation distributions similar to those measured for our subjects, whereas a random strategy does not (C).

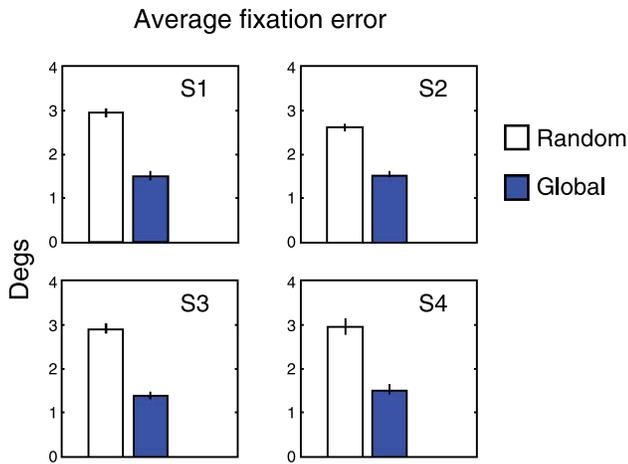


Figure 5. Fixation error. The fixation error between human fixations and the global strategy predictions is significantly smaller than the fixation error of the random strategy.

between humans and different strategies in their compactness, shape, and so forth.

Receiver operating characteristic (ROC)

Using the spatial prediction map, we can conduct a rigorous signal detection theory analysis and investigate how well the global strategy predicts individual fixations. Each new fixation (f) is overlaid on the current strategy map, which is updated using the previous series of fixations ($1, \dots, f - 1$). The map is rescaled from 0 to 1, and the prediction value is taken as the maximum value that falls within 1° of the human fixation (Figure 6A), following the approach outlined by Tatler, Baddeley, and

Gilchrist (2005). A criterion window of 1° allows some wiggle room for natural fixation error and imprecision in our sampling of the global prediction (we interpolate a grid with 0.25° spaced samples). Because it is unlikely that information for eye-movement planning is processed in less than 100 ms (Araujo, Kowler, & Pavel, 2001; Caspi, Beutter, & Eckstein, 2004), the prediction map is only updated by a fixation if its dwell time exceeds this value.

Next, we compute ROC curves and measure the area under the curve (AUC) to assess the power of the global strategy prediction. We compute “hits” as the probability that the prediction value exceeds threshold at fixated locations. We compute “false alarms” as the probability that the prediction value exceeds threshold at locations *not* fixated by the observer. We determine “not-fixated” locations by simply evaluating locations predicted by the random strategy. Hits and false alarms are plotted with changing threshold, sweeping out the ROC curve. If the global prediction is no better than random at predicting human fixations, the ROC curve should lie along the positive diagonal (AUC = 0.5). If the global strategy is a good predictor of human fixations, it will tend toward the upper left-hand corner of the plot (AUC = 1.0). To assess the significance of the AUC, we resampled the hits and false alarms in our ROC analysis with replacement to produce bootstrapped estimates. A prediction is considered significantly better than chance if the 95% confidence interval for the AUC does not include 0.5. Figure 6B shows that for all of our observers, the global model is significantly better than chance at predicting the next fixation.

From the fixation error and ROC results, we might be tempted to conclude that human observers use a global

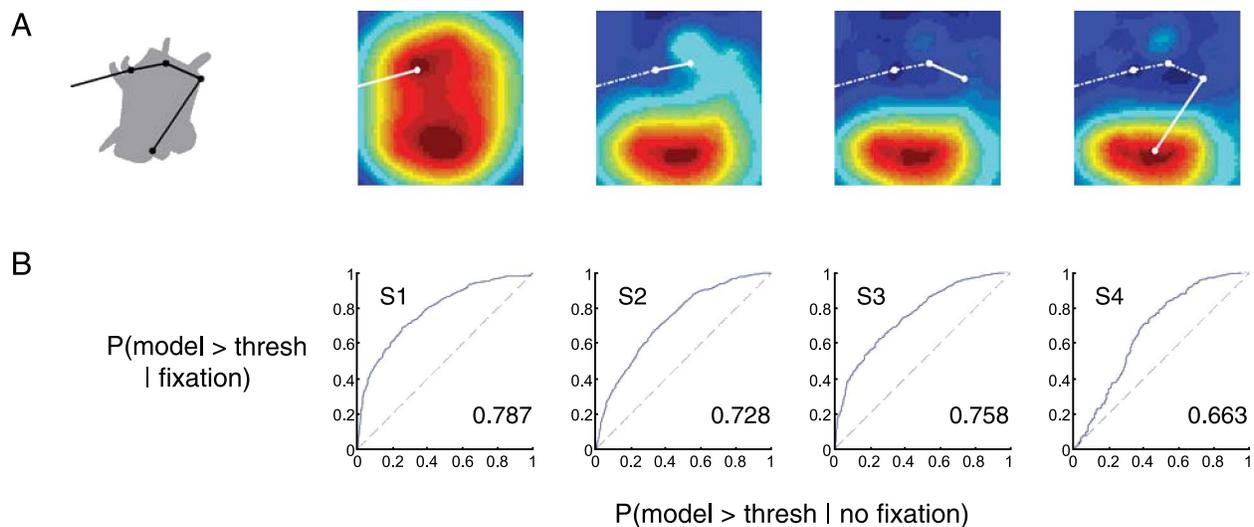


Figure 6. Comparison of human fixation sequence to the global strategy. (A) One observer’s fixation sequence superimposed on a shape (left) and on the corresponding global strategy prediction, which is updated after each fixation. Maps are scaled from 0 to 1 (blue to red). (B) ROC curves for all four observers show that the global strategy is significantly better than the random strategy at predicting fixation locations. The area under the ROC curve is noted on each plot.

information maximization (uncertainty reduction) strategy when planning eye movements to study the novel shape silhouettes in this task. Be aware, however, that these results are generated as a comparison to a *uniform* random model. The simple fact that the global model produces a donut-shaped distribution of fixations may be enough to align it with human fixation patterns. A much more stringent test is one that compares the performance of a strategy against a “smarter” random strategy that knows shape information is near the edges in this task.

The smart random strategy

One way to factor out the bias in the human fixation pattern is to adopt a baseline comparison that has the same

bias (Tatler et al., 2005). To achieve this, we analyze each observer’s fixations on a given trial against his or her own fixations drawn randomly from other trials. The now “smarter” random strategy again provides a measure of what is *not fixated*, from which we generate false alarms for the ROC analysis. Using this much stricter test, how well does the global strategy predict human fixations? Figure 7 demonstrates that the fixation error is lower for the smart random strategy compared with the uniform random strategy, but the global strategy still has a significantly smaller error. When the smart random strategy serves as a baseline comparison, ROC curves shift toward the diagonal but the AUC is still significantly greater than 0.5. The magnitude of the AUC values demonstrate that, although far from perfect, the global strategy has some power to predict human eye movements.

We refer to the global strategy as the “omniscient” strategy because the benefit of all possible fixations is fully known before a decision is made about the best next fixation. We do this by simulating a fixation to a location and computing the information gained. In reality, the visual system cannot possibly compute the global strategy in this manner. More likely, it uses estimates (e.g., heuristics or learned priors) to determine the benefit of each possible next fixation. Such approaches have been taken in the literature (Geman & Jedynak, 1996; Raj et al., 2005). When estimates are used, the global strategy is to maximize the *expected* information rather than the actual information. It is unclear how the visual system would do this without complex computation. Is there a simpler, more efficient strategy that produces similar fixation behavior?

Other strategies

In this section, we consider two biologically plausible strategies for making eye-movement decisions. We evaluate each strategy against the smart random baseline.

Saliency

Given that the shapes in the psychophysical task are novel, top-down influences such as familiarity should be minimized and observers may simply look at salient points on the shape. Locations become salient as their properties (contrast, orientation, motion, etc.) differ from the surrounding locations. In our stimuli, salient locations are those that have an orientation that differs from its surround, such as corners or sharp points. We produced saliency prediction maps for our stimuli (Figure 8A) using the model developed by Itti and Koch (2000), which is available on the web. The version of the saliency model

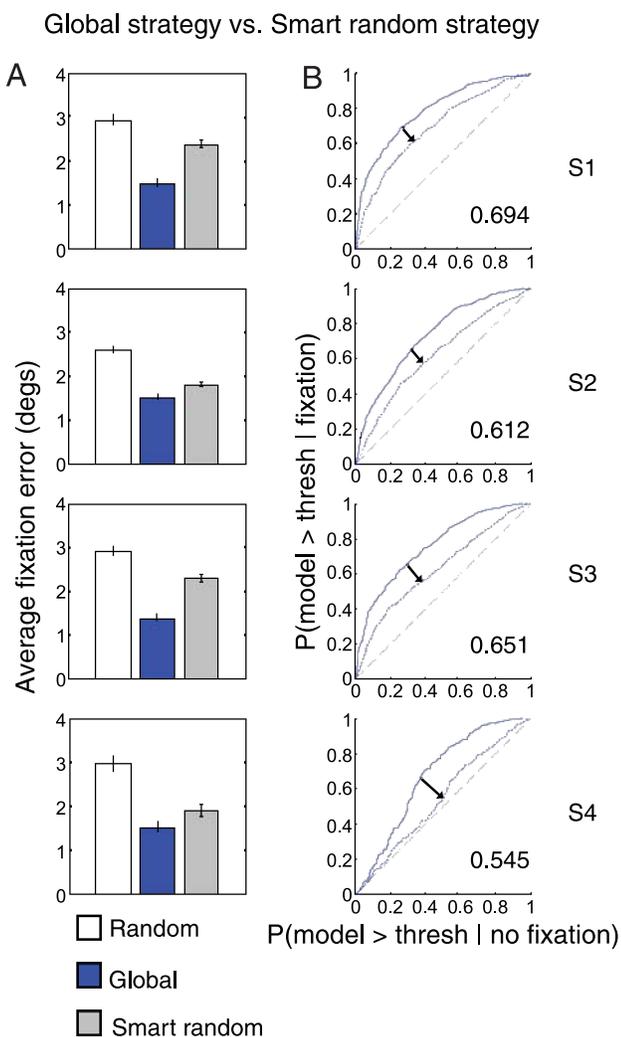


Figure 7. Global strategy versus smart random strategy. (A) Fixation errors of the random, global, and smart random strategies for four subjects. (B) As indicated by the arrows, AUC values are significantly lowered when a smart random model is taken as the baseline comparison. These new values (displayed on each plot) demonstrate that the global strategy still shows significant predictive power.

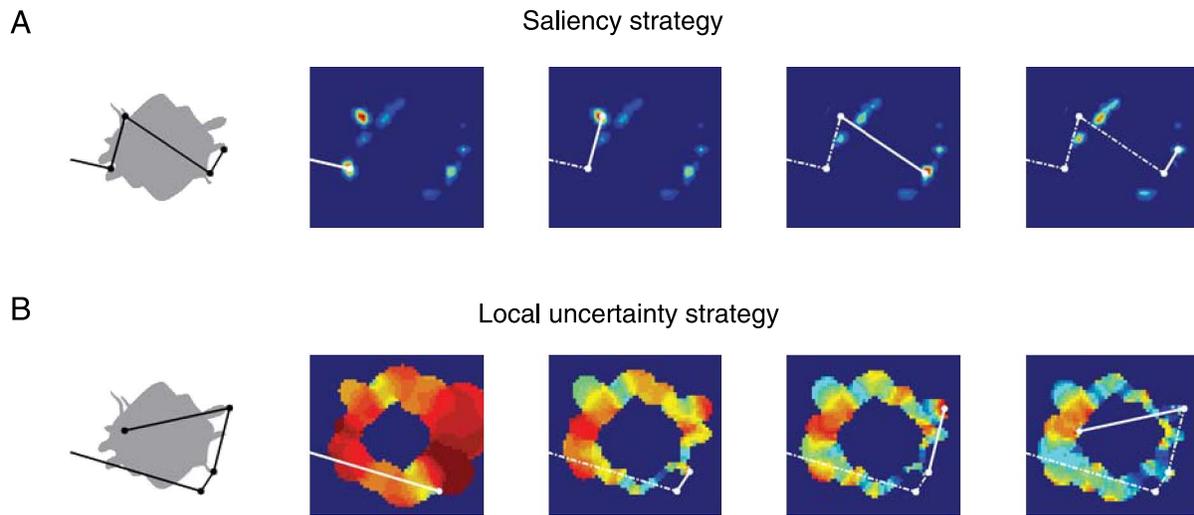


Figure 8. Prediction sequences for saliency and local uncertainty strategies. (A) Saliency: Prediction sequence is displayed on the shape (left). The strategy map for each prediction is shown on the right. Previously predicted locations are blanked out to simulate IOR. (B) Local uncertainty: The prediction map is updated based on the history of human fixations. Maps are scaled 0 to 1 (blue to red).

used here does not take eccentricity factors into account, although newer implementations do. This limitation will be addressed in the discussion.

Local uncertainty

Rather than maximizing information gain globally, the visual system may use a greedy strategy in which only the most informative points, or points of maximum entropy, are fixated. To better understand this difference, imagine two nearby locations that have similar prediction values. The global strategy might be to fixate between them to maximize information about both locations, whereas the local uncertainty strategy would fixate the one with slightly higher uncertainty (more information). To model this, we used the RDE map from Figure 2D directly as a prediction map and the strategy is to fixate the “hot spots” (Figure 8B). This strategy is analogous to the MAP prediction described by Najemnik and Geisler (2005).

Analysis

Both the saliency and local uncertainty strategies produce a donut-shaped distribution, but neither strategy shows a distribution of saccade amplitudes exactly like the observers (Figure 9). As before, we compute the fixation error and ROC curves for these two strategies. In the case of the saliency strategy, we include a 1° mask that inhibits saliency signals at previously fixated locations. We do this to mimic the dynamic changes in the saliency map due to IOR, as in Itti and Koch’s original model. This will presumably improve the prediction of the saliency strategy by reducing the number of salient locations that the

random strategy may predict. For the local uncertainty strategy, the RDE map is updated from the history of human fixations. For both strategies, the prediction strength for the next fixation is evaluated using the maximum value of the strategy prediction map within 1° of the fixation.

Fixation error

Figure 10A plots the fixation error for all strategies. The dashed line represents the error level for the smart random strategy. Neither the saliency nor the local uncertainty

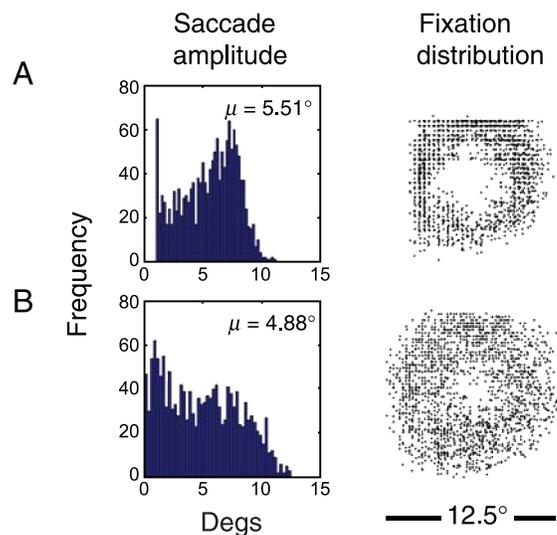


Figure 9. Predicted fixation behavior for saliency and local uncertainty strategies. Saccade amplitude and fixation distributions for the (A) saliency and (B) local uncertainty strategies.

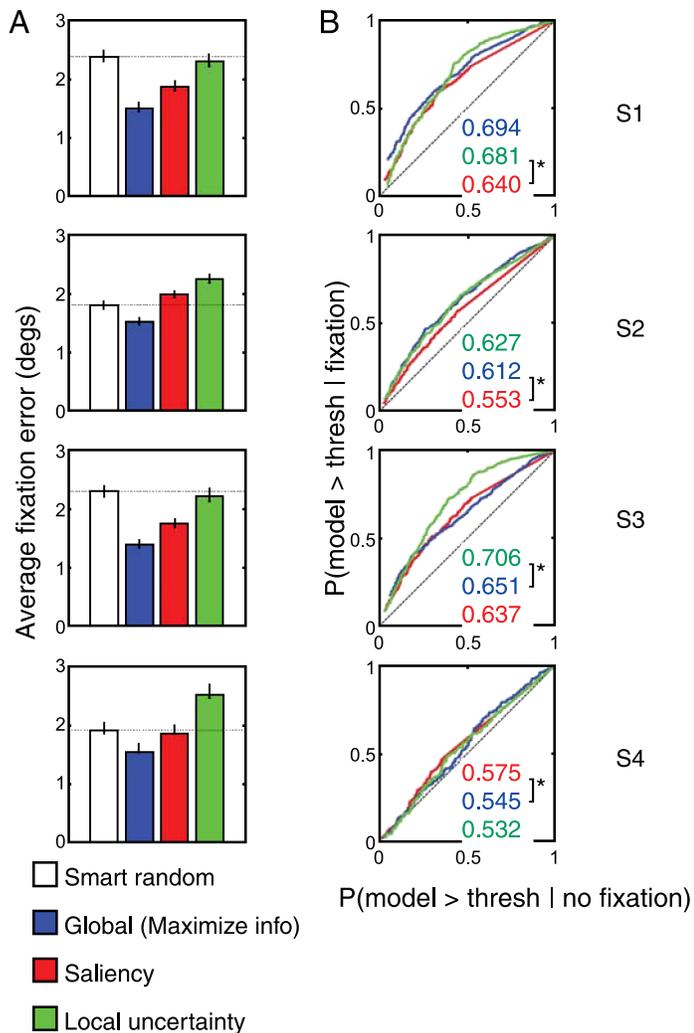


Figure 10. Analysis of optimal, saliency, and local uncertainty strategies. (A) Fixation errors of all strategies, across four observers. (B) ROC curves for the optimal, saliency, and local uncertainty strategies, as compared with the smart random strategy. Brackets indicate significant increases between AUC values.

strategy performs equal to or better than the global strategy according to this measure. In fact, the strategies are sometimes worse than random! Recall that this metric ignores the sequence in which fixations are made and simply computes the distance between observed and predicted fixation locations on a given trial. Both the saliency and local uncertainty strategies have a less compact spatial distribution than our observers, which may explain the larger errors with this metric.

Receiver operating characteristic

Figure 10B plots the ROC curves for the global, saliency, and local uncertainty strategies, as compared with the smart random strategy. The AUC is significantly

greater than 0.5 for all curves across all observers, although it is obvious that Subject 4 is not well predicted by any of the strategies. This is not surprising given that her spatial fixation distribution is very different from the other three subjects; it is highly compact rather than donut shaped.

In Figure 10B, the AUC values are listed in a decreasing-magnitude order and are color coded according to strategy. Brackets indicate a significant difference between values, as determined by bootstrapped 95% confidence intervals. Looking at the data for our first three subjects, we get a very interesting result! Despite the differences in saccade amplitude and fixation distributions, the local uncertainty strategy is at least as good as the global strategy at predicting where observers will look next. The saliency strategy is again a poor predictor. Unlike the fixation error analysis, recall that the ROC analysis does take the sequence of previous human fixations into account and evaluates the ability of each strategy to predict the *next* fixation.

This curious result implies that local uncertainty is providing a valid cue to the eye-movement planning decision. Notice in Figure 9B that the predicted spatial fixation distribution for this strategy is quite diffuse. The discrepancy between fixation error and the ROC finding could be explained if observers consistently undershoot the maximum of the local uncertainty prediction but still land within a hot spot. It is well known that humans make fixations toward the centroids of small shapes (Melcher & Kowler, 1999). What if observers are combining the local uncertainty strategy with a simple centroid prior when planning fixations?

Local uncertainty + centroid

To test this idea, we will assume that observed fixation locations f are biased toward the centroid by a weight w :

$$\begin{bmatrix} f_x \\ f_y \end{bmatrix} = w \begin{bmatrix} C_x \\ C_y \end{bmatrix} + (1-w) \begin{bmatrix} \hat{f}_x \\ \hat{f}_y \end{bmatrix}, \quad (2)$$

where C is the centroid and \hat{f} is the strategy-defined prediction. Figure 11A plots the predicted saccade amplitude and fixation distributions for the local uncertainty strategy with a centroid weighting of 0.25. The mean saccade amplitude is now similar to subject data. The occurrence of many shorter saccades (and resulting bimodal distribution, Figures 9B and 11A) may be an artifact of our local uncertainty computation (see the [General discussion: Local uncertainty](#) section). As expected, the spatial distribution of predicted fixations has a more compact donut shape and looks strikingly similar to the human pattern. This improved distribution is reflected in the decrease in fixation error (Figure 11B).

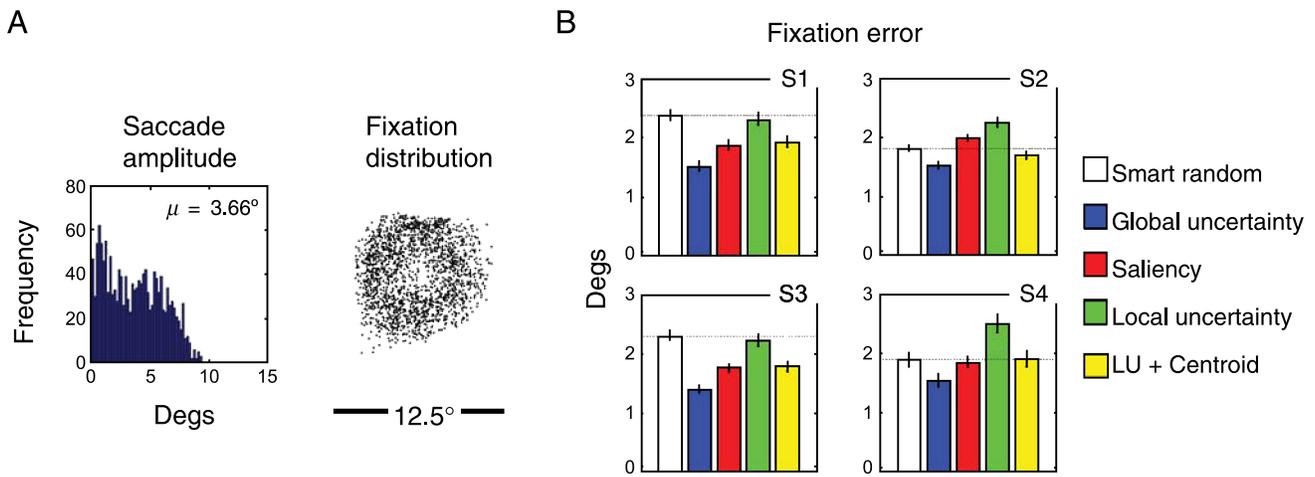


Figure 11. Effect of centroid bias on local uncertainty predictions. (A) Predicted saccade amplitude and fixation distributions for the local uncertainty strategy with a centroid weighting of 0.25. (B) Summary of fixation errors for all strategies.

Given observed fixation locations and different values of w , we can calculate the observer’s *intended* fixation and superimpose it on our local uncertainty strategy map. Using the prediction values from these maps, we again compute ROC curves. Figure 12 plots the AUC as a function of centroid weighting for each subject. The straight lines indicate the baseline AUC for the global and local uncertainty strategies (i.e., without centroid weighting). The 95% confidence intervals attained with bootstrapping allow us to determine which points are significant. For all subjects, the local uncertainty strategy with centroid weighting provides the best prediction of human fixation locations.

General discussion

Much research has gone into understanding the stimulus properties and decision strategies that might underlie human fixation patterns. This article uses an active information-gathering task to mimic situations that are often encountered in natural vision. Here, we discuss several aspects of our approach and findings.

Observer variability

On a given trial, two observers may exhibit very different scan paths. Subject 3 repeated the entire experiment, and her second-pass scan paths were not necessarily similar to her first. Given longer display durations, it is possible that the chosen fixation locations between observers would eventually overlap. This would suggest that there are many fixation sequences that are good enough for collecting the task-relevant information. Using an ROC analysis, there-

fore, provides much better insight into how well different strategies align with human fixation behavior. Some of the variability among subjects may also be attributed to

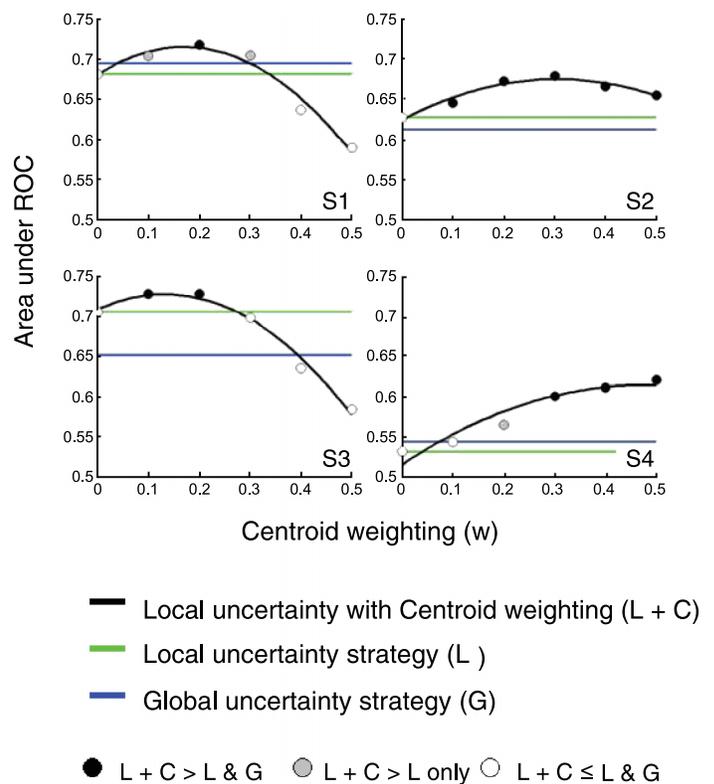


Figure 12. ROC analysis of centroid bias effect. Adding a centroid bias to the local uncertainty prediction results in a significant improvement over other strategies (black symbols). For comparison, the green and blue lines represent the local and global uncertainty strategy predictions. The symbol color indicates significant differences based on 95% confidence intervals.

individual differences in orientation discrimination abilities. Subject 4 had an extremely compact distribution of fixations yet did quite well in the task. It seems that she was able to use more peripheral vision to gather the same quality of information as the other subjects. We chose to use vernier acuity parameters to model orientation discrimination for all of our subjects. Aside from possible individual differences, these parameters only provide an approximation for how the visual system is able to characterize orientation along the bounding contour of a shape. Inhibition from crowding and facilitation due to contour continuation are likely factors that will affect how well contour orientations can be discriminated.

ROC analysis

Despite our choice of approximating parameters, we are able to do a reasonably good job of predicting human fixation locations within an information-theory framework. By using ROC analyses, we get a quantitative and meaningful measure of how well hypothesized strategies predict human fixation locations. If we assume a uniform random baseline, all strategies do an excellent job of predicting fixations because they all gravitate toward object edges—where the stimulus information is and where humans look. If we instead use a “smart” random strategy as our baseline measure to factor out this bias, we can better gauge the power of different strategies to predict individual fixations *within* the overall fixation distribution. This stringent test allows us to compare the microstructure of different strategies and to better discriminate between them.

Saliency

The idea that we look at salient locations has received much attention in the literature. In our ROC analysis, we find that saliency has the least power to predict human fixation locations. We were not particularly surprised by this finding. Our task requires that observers actively gather orientation information along the entire contour of the object. All contour information is important to the task, not just the salient information (e.g., corners). In this active setting, information-based models of eye movements may do a better job. It should be noted, however, that we implemented the classic saliency model of Itti and Koch (2000). A saliency model that factors in eccentricity effects may perform differently. The implementation of Peters, Iyer, Itti, and Koch (2005) attenuates orientation and contrast signals as a function of eccentricity before producing a salience map. This leads to shorter saccade amplitude predictions because distant points become less salient. The underlying topology of salience is not affected, however; thus, we speculate that adding eccentricity factors may alter the sequence of predicted fixations

without altering their location. Our measure of fixation error is likely accurate then, and the saliency strategy demonstrates larger errors than the global information strategy. Because eccentric salient points are attenuated, the random strategy in the ROC analysis would hit fewer predicted locations, reducing false alarms. Human fixations would also have less chance of landing on salient locations, reducing hits. Thus, it is unclear how the AUC would be affected by adding eccentricity factors.

Even without eccentricity factors, our analysis does show some weak predictive power for the saliency strategy, but the results may be somewhat confounded. In our stimuli, local uncertainty and saliency predictions often overlap, especially early in the fixation sequence. This correlation is likely present in all natural stimuli. Stimuli that cleanly isolate local uncertainty and saliency effects would be needed to determine if the visual system makes use of only one strategy or if it uses both strategies.

Maximize information (global)

Consistent with the findings in visual search (Najemnik & Geisler, 2005), we find that observer fixation placement is well described by an optimal strategy that seeks to maximize information gain across the stimulus. The computation of this prediction is intensive and requires an exhaustive evaluation of all possible fixations before a decision is made. We compute the prediction by actually simulating each possible next fixation and computing the information gained. The visual system might achieve a similar computation by using heuristics or estimates of orientation information (Raj et al., 2005). This *expected* information gain must still be computed globally across the visual field.

Local uncertainty

A simpler strategy would be to just “look at,” that is, foveate the region about which we are most uncertain (i.e., the most locally informative point). Our ROC analysis revealed that despite its more diffuse predicted fixation distribution, the local uncertainty strategy does as good a job of predicting fixation locations as the global strategy. This suggests that the local uncertainty signal is powerful. However, our prediction map may not be correct in detail. Our inference of orientation at a point, discretization into eight bins, and use of vernier parameters are all approximations that may introduce error into our estimation of local uncertainty. Isolated maxima will predict a fixation regardless of neighboring activity. This may be the underlying cause of the bimodal distribution (Figure 9B) of saccade length for this strategy—often shorter saccades are made to isolated maxima (see Figure 8B, third panel). The visual system, perhaps through lateral interactions,

may smooth these spurious signals. Also, large areas of activity may be sharpened through nonlinear competition. Such manipulations to the prediction map are easily explored, but it is probably better to do so after the fundamental parameters for contour processing as a function of eccentricity have been better quantified.

Saccade length

Human saccade amplitudes have been reported to follow a characteristic distribution with a mode around 3° and amplitudes rarely exceeding 15° (Bahill et al., 1975). Data from our subjects are consistent with this report; however, the three aforementioned strategies predict a mode considerably larger than 3° . This discrepancy could be due to an incorrect estimation of the visibility of contours in the periphery (i.e., uncertainty in the periphery is even greater, drawing predictions further out), or it may be explained by factors that we have not considered. To our knowledge, there are no mechanical factors that restrict the length of a saccade, but it is possible that energy constraints favor shorter saccades or that time pressure leads observers to make shorter saccades. When a centroid bias is added to the local uncertainty strategy, the mode of saccade amplitudes falls in closer alignment with human data, and the prediction of fixation placement is greatly improved.

Biological plausibility

Lee and Yu (2000) proposed that a local uncertainty or information signal may underlie eye-movement planning and could be encoded as early as orientation hypercolumns in visual area V1. Each hypercolumn represents activity over different orientations for a given location in the visual field, analogous to our histogram representation. Lateral connections, feedback connections, or both could serve to reach a local consensus about orientation content, smoothing the map and reducing the spurious signals we see in our prediction maps (Hamker, 2003; Lee & Mumford, 2003). This early uncertainty map would then need to be combined with a stimulus-centered representation that incorporates knowledge gained from previous fixations, possibly mediated by mechanisms that pool orientation at a single scale (Olzak & Thomas, 1992). This speculative architecture for eye-movement planning resonates well with the theories of redundancy reduction that have been proposed as a guiding principle for the evolution of our visual system.

Framework-driven considerations

Regardless of how or where in the brain eye-movement decisions are made, the information-theory framework

provides an invaluable tool for outlining a principled approach to research on this topic. In the work presented here, we have made several assumptions out of necessity, and these assumptions must be challenged! For example, we have assumed that shape knowledge is updated before the next fixation is planned. Several studies suggest that rather than being programmed one at a time, a sequence of fixations may be planned at once (Caspi et al., 2004; McPeck, Skavenski, & Nakayama, 2000). Is there a fixed integration time for including new information into the eye-movement plan, or does it depend on how much information is available during a given fixation? Is the integration time altered under different stimulus or task conditions?

When updating information, we assume that orientations along the contour are independent, although real-world objects are constrained to have a closed, piecewise smooth contour. Observers may use this prior world knowledge to make inferences about contour orientation.

Finally, we have assumed a noiseless system, whereas work in neurophysiology suggests that noise may affect which of two competing locations or decisions will be selected (Carpenter, 1988; Shadlen & Newsome, 2001). Internal noise may also degrade the information signal over time, whereas we have assumed perfect memory. Studying subject self-variability in a repeated eye-movement decision task may shed some light on the nature of decision noise.

Our results suggest that a local uncertainty rule may dominate eye-movement decisions. In our experiment, the fixation distribution data suggest that two tasks were being performed: (1) localizing the shape and then (2) learning the boundary. We focused our analysis on the second task. Questions still remain as to whether different strategies may be used in different tasks or if there is a switching between strategies during a single task. We hypothesize that the decision strategy remains fixed but that the *task-relevant information* changes. Defining and quantifying information for a variety of tasks remains one of the great challenges of vision research.

Summary and conclusion

Information theory provides an elegant framework for conceptualizing and modeling human eye-movement behavior. To model orientation information in our task, we selected vernier acuity parameters from the literature; however, further research is needed to characterize how orientation information along a continuous contour is processed by the visual system. Using signal detection theory, we rigorously compared the predictions of several decision strategies for predicting fixation locations: maximize total information (global), saliency, and local uncertainty. Saliency was a poor predictor of

fixation placement in our active, information-seeking task. The global strategy provided a good fit to the observer data. It is not clear whether the visual system could compute this decision without use of estimates or heuristics. Local uncertainty also provided a good fit to the data and could easily be computed as early as V1 orientation hypercolumns. Combining this decision with a simple centroid weighting provided the best fit to human fixations, suggesting that other factors may play a role into making the final decision on where to move the eyes. Further research is needed to determine if different strategies are used under different conditions or whether observers are able to use hybrid strategies.

Appendix A: Model details

Probabilistic model

Our psychophysical task entails learning the shape of a novel silhouette. The brief display time makes the task challenging because the visual information obtained from one fixation is insufficient to determine the shape exactly, given the reduced resolution of the periphery. In this [Appendix](#), we describe a probabilistic model for representing visual information about the stimulus and how this representation is updated with information acquired from new fixations.

We represent the stimulus shape as a collection of *edgelets* or small straight-line segments that approximate the continuous shape boundary. Each edgelet can assume any one of eight possible orientations, which is a discretization of all possible orientations from 0° to 180°. There are a total of n edgelet orientations along the boundary, labeled x_i , where $i = 1, 2, \dots, n$, and $x_i = 1, 2, \dots, 8$ for each i . We set n to be equal to the number of boundary or edge pixels. The edgelet orientations are unknown to the observer and need to be inferred from visual information.

We have defined this edgelet representation rather than a pixel-based representation to reduce our computational load. This simplification incorrectly assumes perfect knowledge of edge locations in the stimulus, completely ignoring positional uncertainty. However, positional uncertainty is roughly 10-fold less than orientation uncertainty across eccentricities (Levi et al. 1985; White, Levi, & Aitsebaomo, 1992). Thus, ignoring positional uncertainty is unlikely to affect the topology of our strategy prediction maps.

The visual information obtained about the edgelets is modeled using the responses of a bank of filters that measure the frequency of each orientation in a local region. The responses of a population of oriented filters within a neighborhood of radius $r(E)$ are represented as a

histogram over eight orientations. We choose $r(E)$ to be equal in size to a “perceptive hypercolumn,” as described by Levi et al. (1985) for vernier acuity in the periphery. Specifically, $r(E)$ is the distance at which small flankers begin to elevate thresholds for a vernier acuity stimulus. It is thought that these flankers encroach on the orientation-selective cells that are analyzing the vernier stimulus and is, therefore, a rough measure of orientation hypercolumns. As this is a perceptual finding, Levi et al. coin it the perceptive hypercolumn. Quantitatively,

$$\begin{aligned} r(E) &= s(E + E_2), \\ s &= 0.1, \\ E_2 &= 0.8, \end{aligned} \tag{A1}$$

where E_2 is the eccentricity at which acuity drops to half its value in the fovea and s is the slope. We further interpret $r(E)$ as an effective radius over which the visual system spatially pools orientation information ([Figure A1](#)). Unpublished data from our laboratory support the Levi et al. parameters.

More precisely, let $E_i(\mathbf{F})$ denote the eccentricity of location i relative to fixation \mathbf{F} . Thus, we write $r(E_i(\mathbf{F}))$ for the radius of the histogram at edgelet i given fixation \mathbf{F} . The histogram is normalized by the total number of edgelets within the radius so that all the histogram entries sum to 1. For each edgelet i viewed from fixation \mathbf{F} , we denote the histogram by $\mathbf{h}_i(\mathbf{F})$, where the boldface indicates that it is a vector with eight components (see [Figure A1](#)).

The histogram $\mathbf{h}_i(\mathbf{F})$ provides a summary of the shape boundary near edgelet i . If the boundary is perfectly straight within the receptive field radius, then the histogram will show the presence of only one orientation in the entire local population, which uniquely determines all the edgelet orientations in that population. Conversely, a *flatter*, higher entropy histogram indicates that the local shape is more complex.

We model the evidence that $\mathbf{h}_i(\mathbf{F})$ provides about edgelet orientation x_i using a simple likelihood model: $P(\mathbf{h}_i(\mathbf{F})|x_i, E_i(\mathbf{F})) = h_{i,x_i}(\mathbf{F})/Z$, where $h_{i,x_i}(\mathbf{F})$ is the x_i th component of $\mathbf{h}_i(\mathbf{F})$, that is, the fraction of edgelets within the pooling neighborhood with orientation x_i . Z is a normalization constant. For an intuitive interpretation of the likelihood function, notice that if $\mathbf{h}_i(\mathbf{F})$ is 0 for some component $x_i = z$, then no edgelet in the local population has orientation z ; thus, the likelihood function $P(\mathbf{h}_i(\mathbf{F})|x_i, E_i(\mathbf{F}))$ equals 0 for $x_i = z$, which rules out the possibility that $x_i = z$. Conversely, the higher the value of $P(\mathbf{h}_i(\mathbf{F})|x_i, E_i(\mathbf{F}))$ for any component value $x_i = z$, the more likely that the true value of x_i is actually z .

A simple uniform prior is placed on the distribution of orientations: $P(x_i) = 1/8$, which means that all orientations

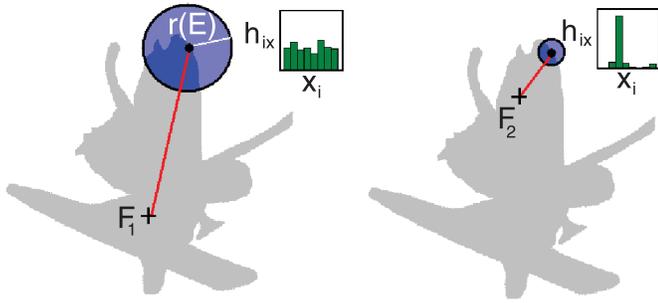


Figure A1. Orientation pooling depends on eccentricity. The size of a pooling neighborhood $r(E)$ depends on the eccentricity E or distance from the current fixation. The left and right panels show local orientation histograms at the same location, for two different fixations.

are a priori equally likely. Using Bayes' rule, we obtain the posterior distribution

$$P(x_i | \mathbf{h}_i(\mathbf{F}), \mathbf{E}_i(\mathbf{F})) = h_{i,x_i}(\mathbf{F}) / Z', \quad (\text{A2})$$

where Z' is a normalization constant.

Given the posterior distributions over all the edgelets, the entropy map is defined as described in the next section. Similarly, the posterior probability can be updated for multiple fixations \mathbf{F}_1 and \mathbf{F}_2 as follows:

$$P(x_i | \mathbf{h}_i(\mathbf{F}_1), \mathbf{h}_i(\mathbf{F}_2), \mathbf{E}_i(\mathbf{F}_1), \mathbf{E}_i(\mathbf{F}_2)) = h_{i,x_i}(\mathbf{F}_1) h_{i,x_i}(\mathbf{F}_2) / Z_2, \quad (\text{A3})$$

where Z_2 is a normalization constant. Although this approximation has some undesirable properties (such as making the marginal distribution more peaked if the same fixation is made repeatedly), it provides a simple mechanism for combining histogram evidence from multiple, distinct fixations.

RDE map

The local uncertainty strategy posits that human subjects plan their next fixation to a location of maximum local uncertainty in the image. Because our edgelet model explicitly represents orientation uncertainty only along the silhouette border, we must extrapolate this representation of uncertainty to all possible locations in the image. We call this extrapolation the RDE map.

The RDE map quantifies the amount of uncertainty at each location in the image, given the resolution falloff relative to the current fixation. The RDE value at pixel i is the sum of entropies corresponding to all edgelets within radius $r(E_i(\mathbf{F}))$ of pixel i and is equal to 0 if there are no edgelets within this radius. In other words, the RDE is

defined at every pixel location i (given fixation \mathbf{F}) as follows:

$$\text{RDE}_i = \sum_{j \in \text{all edgelet locations within radius } r(E_i(\mathbf{F})) \text{ of } i} H_j. \quad (\text{A4})$$

Here, H_j is the entropy of edgelet j ; that is,

$$H_j = - \sum_{z=1}^8 P(x_j = z) \log P(x_j = z), \quad (\text{A5})$$

where we omit the explicit conditioning of posterior probabilities $P(x_j)$ on histogram data (as in Equations A2 and A3) for simplicity.

Future work will build the image representation based on a retinotopic grid of V1-like filters, rather than using the simple edgelet-based representation, which, for simplicity, assumes exact prior knowledge of edge positions but no prior knowledge of edge orientations. In this case, the RDE map will be a more fundamental construct that is tied to our uncertainty across the entire image, not only along the silhouette border.

Fixation prediction strategies

We define two possible strategies for predicting eye movements based on our probabilistic model: one global and one local.

Maximize information (global)

The global strategy predicts the next fixation location to be the one that maximizes information (i.e., reduces global uncertainty). This strategy evaluates fixations to every possible location in the image and then chooses the fixation location that minimizes the total edgelet entropy as the prediction. More precisely, choose fixation location \mathbf{F} to minimize the total entropy of all n edgelets:

$$H_{\text{tot}} = \sum_{i=1}^n H_i. \quad (\text{A6})$$

Of course, this strategy is biologically implausible because it assumes full knowledge of high-resolution (i.e., foveal) image information *everywhere* in the image! The visual system would, instead, need to compute the *expected* information gain through the use of priors or heuristics.

Local uncertainty

The local uncertainty model predicts the next fixation location to be the location of maximal local uncertainty, as

defined by the RDE map. In other words, choose fixation location F to be the pixel location i that is the maximum of RDE_i . Because the RDE map is updated with each new fixation, it is straightforward in determining its maximum, and thus, this strategy could be implemented easily in the human visual system.

Acknowledgments

This research was supported by grants from Smith–Kettlewell and Ruth L. Kirchstein NRSA (#EY 14536-02) to L.W.R.; Air Force (#FA9550-05-1-0151) and NSF (#0347051) to P.V.; and NIDRR (#H133G030080), NSF (#IIS0415310), and NIH (#EY015187-01A2) to J.C.

Commercial relationships: none.

Corresponding author: Laura Walker Renninger.

Email: laura@ski.org.

Address: The Smith–Kettlewell Eye Research Institute, 2318 Fillmore Street, San Francisco, CA 94115, USA.

References

- Araujo, C., Kowler, E., & Pavel, M. (2001). Eye movements during visual search: The costs of choosing the optimal path. *Vision Research*, *41*, 3613–3625. [PubMed]
- Bahill, A. T., Adler, D., & Stark, L. (1975). Most naturally occurring human saccades have magnitudes of 15 degrees or less. *Investigative Ophthalmology*, *14*, 468–469. [PubMed]
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *2*, 433–436. [PubMed]
- Carpenter, R. H. S. (1988). *Movements of the eyes* (2nd ed.). London: Pion.
- Caspi, A., Beutter, B. R., & Eckstein, M. P. (2004). The time course of visual information accrual guiding eye movement decisions. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 13086–13090. [PubMed] [Article]
- Dorris, M. C., Klein, R. M., Everling, S., & Munoz, D. P. (2002). Contribution of the primate superior colliculus to inhibition of return. *Journal of Cognitive Neuroscience*, *14*, 1256–1263. [PubMed]
- Geman, D., & Jedynek, B. (1996). An active testing model for tracking roads from satellite images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *18*, 1–14.
- Hamker, F. H. (2003). The reentry hypothesis: Linking eye movements to visual perception. *Journal of Vision*, *3*(11):14, 808–816, <http://journalofvision.org/3/11/14/>, doi:10.1167/3.11.14. [PubMed] [Article]
- Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*, 188–194. [PubMed]
- Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, *3*(1):6, 49–63, <http://journalofvision.org/3/1/6/>, doi:10.1167/3.1.6. [PubMed] [Article]
- Itti, L., & Baldi, P. (2005). A principled approach to detecting surprising events in video. *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, *1*, 631–637.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*, 1489–1506. [PubMed]
- Land, M., Mennie, N., & Rusted, J. (1999) The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*, 1311–1328. [PubMed]
- Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, *20*, 1434–1448. [PubMed]
- Lee, T. S., & Yu, S. (2000). An information-theoretic framework for understanding saccadic eye movements. *Advances in Neural Information Processing Systems*, *12*, 834–840. [Article]
- Legge, G. E., Hooven, T. A., Klitz, T. S., Stephen Mansfield, J. S., & Tjan, B. S. (2002). Mr. Chips 2002: New insights from an ideal-observer model of reading. *Vision Research*, *42*, 2219–2234. [PubMed]
- Legge, G. E., Klitz, T. S., & Tjan, B. S. (1997). Mr. Chips: An ideal-observer model of reading. *Psychological Review*, *104*, 524–553. [PubMed]
- Levi, D. M., Klein, S. A., & Aitsebaomo, A. P. (1985). Vernier acuity, crowding and cortical magnification. *Vision Research*, *25*, 963–977. [PubMed]
- McPeck, R. M., Skavenski, A. A., & Nakayama, K. (2000). Concurrent processing of saccades in visual search. *Vision Research*, *40*, 2499–2516. [PubMed]
- Melcher, D., & Kowler, E. (1999). Shapes, surfaces and saccades. *Vision Research*, *39*, 2929–2946. [PubMed]
- Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, *434*, 387–391. [PubMed]
- Olzak, L. A., & Thomas, J. P. (1992). Configural effects constrain Fourier models of pattern discrimination. *Vision Research*, *32*, 1885–1898. [PubMed]

- Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research*, *45*, 2397–2416. [PubMed]
- Posner, M. I., & Cohen, Y. (1984). In H. Boumas & Y. Bouwhuis (Eds.), *Attention and performance* (vol. X, pp. 531–556). Erlbaum: Hillsdale, NJ.
- Raj, R., Geisler, W. S., Frazor, R. A., & Bovik, A. C. (2005). Contrast statistics for foveated visual systems: Fixation selection by minimizing contrast entropy. *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, *22*, 2039–2049. [PubMed]
- Rao, R. P., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Research*, *42*, 1447–1463. [PubMed]
- Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network*, *10*, 341–350. [PubMed]
- Renninger, L. W., Coughlan, J., Verghese, P., & Malik, J. (2005). An information maximization model of eye movements. *Advances in Neural Information Processing Systems*, *17*, 1121–1128. [PubMed] [Article]
- Renninger, L. W., Verghese, P., & Coughlan, J. (2005a). *Eye movements can be understood within an information theoretic framework*. Computational & Systems Neuroscience (Cosyne05), Salt Lake City, UT.
- Renninger, L. W., Verghese, P., & Coughlan, J. (2005b). Modeling eye movements in a shape discrimination task [Abstract]. *Journal of Vision*, *5*(8):921, 921a, <http://journalofvision.org/5/8/921/>, doi:10.1167/5.8.921.
- Renninger, L. W., Verghese, P., & Coughlan, J. (2006). Do eye movements incorporate knowledge of part structure? [Abstract]. *Journal of Vision*, *6*(6):482, 482a, <http://journalofvision.org/6/6/482/>, doi:10.1167/6.6.482.
- Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, *86*, 1916–1936. [PubMed] [Article]
- Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215. [PubMed]
- Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual determinants of eye movements: Effects of scale and time. *Vision Research*, *45*, 643–659. [PubMed]
- Vishwanath, D., & Kowler, E. (2003). Localization of shapes: Eye movements and perception compared. *Vision Research*, *43*, 1637–1653. [PubMed]
- White, J. M., Levi, D. M., & Aitsebaomo, A. P. (1992). Spatial localization without visual references. *Vision Research*, *32*, 513–526. [PubMed]
- Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.