

Probabilistic modeling of eye movement data during conjunction search via feature-based attention

Ueli Rutishauser

Computation and Neural Systems,
California Institute of Technology,
Pasadena, CA, USA



Christof Koch

Division of Biology and Division of Engineering and
Applied Science, California Institute of Technology,
Pasadena, CA, USA



Where the eyes fixate during search is not random; rather, gaze reflects the combination of information about the target and the visual input. It is not clear, however, what information about a target is used to bias the underlying neuronal responses. We here engage subjects in a variety of simple conjunction search tasks while tracking their eye movements. We derive a generative model that reproduces these eye movements and calculate the conditional probabilities that observers fixate, given the target, on or near an item in the display sharing a specific feature with the target. We use these probabilities to infer which features were biased by top-down attention: Color seems to be the dominant stimulus dimension for guiding search, followed by object size, and lastly orientation. We use the number of fixations it took to find the target as a measure of task difficulty. We find that only a model that biases multiple feature dimensions in a hierarchical manner can account for the data. Contrary to common assumptions, memory plays almost no role in search performance. Our model can be fit to average data of multiple subjects or to individual subjects. Small variations of a few key parameters account well for the intersubject differences. The model is compatible with neurophysiological findings of V4 and frontal eye fields (FEF) neurons and predicts the gain modulation of these cells.

Keywords: visual attention, eye movements, computational model, probabilistic, visual search, V4, FEF

Citation: Rutishauser, U., & Koch, C. (2007). Probabilistic modeling of eye movement data during conjunction search via feature-based attention. *Journal of Vision*, 7(6):5, 1–20, <http://journalofvision.org/7/6/5/>, doi:10.1167/7.6.5.

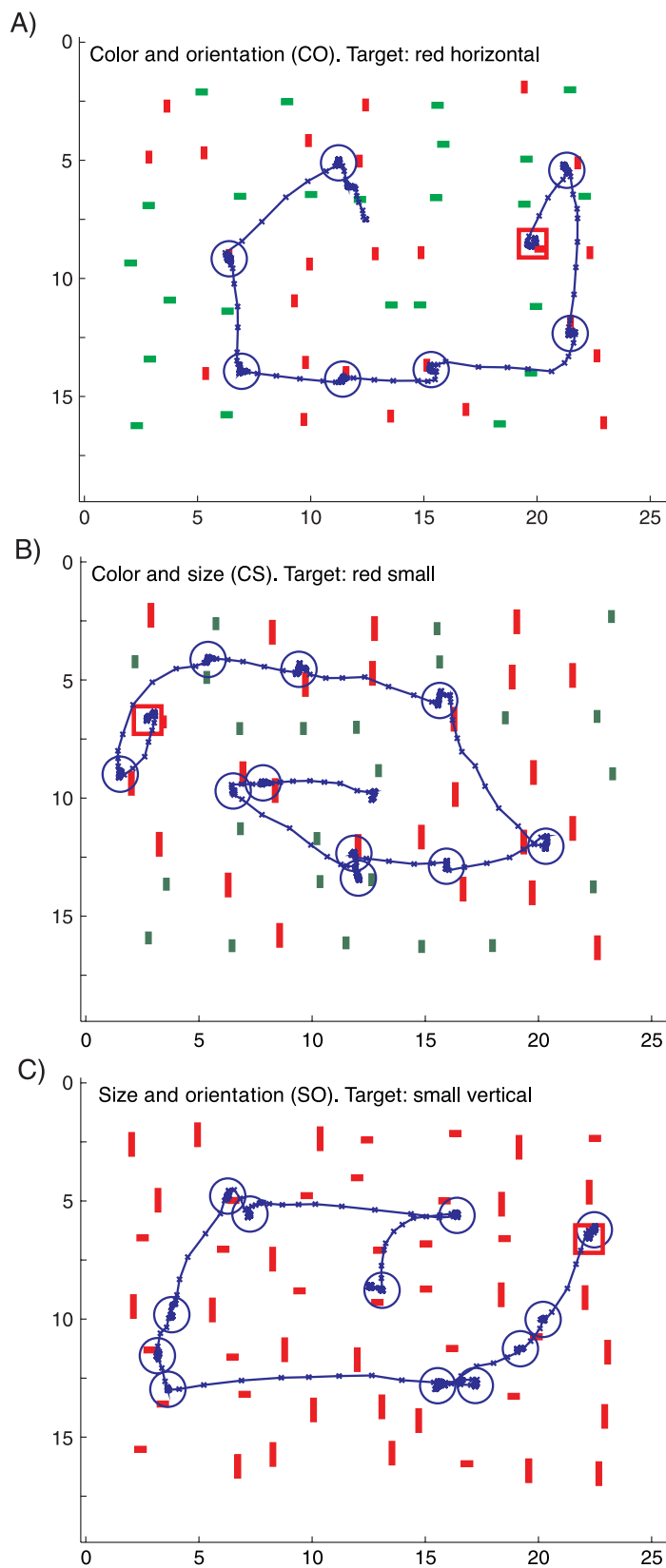
Introduction

When we are looking for a known item in a cluttered scene, our eye movements are not random. Rather, the path our eyes travel on during search reflects an estimate of where the item is most likely located. How these locations are chosen by the visual system remains poorly understood. The visual system needs to combine the retinal input with knowledge about the target in some way to estimate the most probable target location. One possibility would be to disregard the information about the target and to only use the visual input to guide search (a pure bottom-up strategy as in saliency-based models of attention; Itti & Koch, 2001). Clearly, this strategy is not appropriate for search—depending on the task, different locations are fixated. The other possibility is to combine the incoming image with information about the target, that is, top-down information modulating a bottom-up strategy (Navalpakkam & Itti, 2005; Rao & Ballard, 1997; Tsotsos et al., 1995; Wolfe, 1994). This is the approach we take here. For any such strategy, a crucial decision is which feature(s) of the target guides the search. There are many theoretical models (Najemnik & Geisler, 2005; Navalpakkam & Itti, 2005; Rao, Zelinsky, Hayhoe, &

Ballard, 2002; Treisman & Gelade, 1980; Verghese, 2001; Wolfe, 1994) of which features the visual system should use; however, this leaves open the question of which features are actually used. Here, we provide a quantitative estimate of which features guide search (e.g., if the target is defined by multiple features, such as color and orientation, which ones are used to bias search). Note that in this paper, *top-down attention* refers to the mechanism of selectively biasing the weights of specific feature channels regardless of spatial location (Desimone & Duncan, 1995; Itti & Koch, 2001; Muller, Heller, & Ziegler, 1995; Palmer, 1994). We construct a computational model in close analogy to the neuronal processes underlying saccade planning that generates realistic eye movements and compare these against our own data.

We used conjunction search arrays (Figure 1) that contain 49 elements, one of which was the target. Except for some control experiments, the target was uniquely defined by two features. The 48 distractors were chosen such that one half (24) shares one of the features of the target, whereas the other half shares the other. This choice of distractors assures that potential biases in subselecting items (e.g., all of same color vs. same orientation) do not change the number of elements that need to be processed. We used three feature dimensions (color, size, and

orientation) with two variations each (red/green, big/small, and horizontal/vertical). In any task, only two of the feature dimensions were varied whereas the other was kept constant. This yields three different task conditions:



color and orientation (CO, Figure 1A), color and size (CS, Figure 1B), and size and orientation (SO, Figure 1C).

Visual search paradigms are widely used for both behavioral experiments in humans (Palmer, Verghese, & Pavel, 2000; Wolfe, 1998) as well as electrophysiological experiments in nonhuman primates (Bichot, Rossi, & Desimone, 2005; Chelazzi, Miller, Duncan, & Desimone, 1993; Motter, 1994; Motter & Belky, 1998; Ogawa & Komatsu, 2006). It is thus of great importance to understand the underlying neuronal mechanisms of this process. Information about the target (location or identity) improves search performance and detection accuracy (Burgess & Ghandeharian, 1984; Muller et al., 1995; Swenson & Judy, 1981) and therefore can be utilized to structure the search process. How this a priori information about the target is used, however, remains unclear. How knowledge of a particular feature benefits search performance has been traditionally measured by how long it takes to find a target among a varying number of distractors (Treisman & Gelade, 1980; Wolfe, 1994). Although such reaction time (RT) measurements as a function of set size allow quantification of search efficiency (Treisman & Gelade, 1980; Wolfe & Horowitz, 2004), it remains unclear why certain features are inefficient and why others are not. Similarly, RT measurements have demonstrated that knowledge about a target defined by multiple features improves search performance (conjunction search). However, it is not clear how this improvement in performance is achieved. Imagine, for example, an artificial search array where items are defined by orientation and color only. Behavioral experiments indicate that monkeys that search for a unique target in such a display selectively fixate near items that have the same color as the target (Motter & Belky, 1998). This occurs despite the fact that the orientation of the target was shared by equally many distractors and is equally easy to distinguish. This implies the monkey made a choice, explicit or implicit, to use color as the guiding feature. Does this mean that the other piece of information about the target, the orientation, was not used to guide the search? Here we will show that both are used, but differently.

It is well known that activity in the frontal eye fields (FEF) is sufficient to trigger eye movements (Bruce & Goldberg, 1985; Schall, Hanes, Thompson, & King, 1995). This is demonstrated convincingly by microstimulation (Bruce, Goldberg, Bushnell, & Stanton, 1985). Some V4 neurons respond selectively to the color,

Figure 1. Examples of search paths in three different conditions. Shown are all data points recorded at 250 Hz from three different subjects. Fixations are indicated with a circle or a rectangle (last fixation) and always start out at the center. (A) Color and orientation (CO) condition, the target is red horizontal. Note the color bias. (B) Color and size (CS), the target is red small. Note the color bias. (C) Size and orientation (SO) condition, the target is small vertical. Note the size bias. The units of both x and y axes are in degrees of visual angle as seen by the subject.

orientation, and shape of visual stimuli (Gallant, Connor, Rakshit, Lewis, & Van Essen, 1996; McAdams & Maunsell, 1999a; Motter, 1994). For the same physical input and eye location, these feature selective neurons respond differently based on top-down information (Bichot et al., 2005; Motter, 1994). V4 projects to both inferior temporal cortex (IT) as well as FEF (Schall, Morel, King, & Bullier, 1995). The opposite is also true: FEF neurons project directly into V4 (Stanton, Bruce, & Goldberg, 1995). This allows direct modulation of V4 activity by microstimulation of the FEF (Moore & Armstrong, 2003).

The firing activity of FEF neurons needs to reach a particular threshold until an eye movement is initiated (“race to threshold”). Using this mechanism, FEF neurons are thought to accumulate evidence in favor of initiating a saccade (Hanes & Schall, 1996) or an attentional shift (Moore & Fallah, 2004), to a particular location. It is therefore likely that the locations receiving most input from V4 reach threshold fastest. We point out parallels of this neuronal architecture with our model of saccade generation that allows us to predict how V4 and FEF neurons will respond during visual search.

We first present eye tracking data recorded while subjects performed our search task. Subsequently, we construct a computational model to demonstrate that these eye movements are dominated by two parameters which closely resemble the top-down gain modulation of V4 and FEF neurons. Other parameters commonly implicated in search performance are of much less significance. This is especially true for memory capacity (Horowitz & Wolfe, 1998; McCarley, Wang, Kramer, Irwin, & Peterson, 2003).

Methods

Subjects

Nine subjects (first experiment) and three subjects (second experiment) were paid for participation in the experiment. None of the subjects were aware of the purpose of the experiment. The experiments were approved by the Caltech Institutional Review Board, and all subjects gave written informed consent.

Experimental setup

Eye movements (Experiment 1) were recorded with a headmounted EyeLinkII (SR Research, Canada) system. We recorded binocularly (250 Hz sampling) but only used the information of the more reliable eye. Experiments were implemented in Matlab using the psychophysics and EyeLink Toolbox extensions (Brainard, 1997; Cornelissen, Peters, & Palmer, 2002). Subjects were seated 80 cm in

front of the screen using a chin rest to assure minimal head movements. The search display area was $25^\circ \times 20^\circ$. The only manual interactions with the experimental system were by pressing a button on a gamepad. For calibration, the built-in 9-point calibration grid was used. Calibration was repeated as necessary during the experiment. The effective radial resolution was 0.6° after calibration. Fixation locations were obtained with the built-in fixation detection mechanism. Each subject typically performed 2 blocks of the same condition and 8–12 blocks in total (see below). Thus, each subject performed only a subset of all experimental conditions (of which there were 8).

We recorded additional control subjects (Experiment 2) on an EyeLink 1000 System (SR Research, Canada). We recorded movements of the right eye with 1,000 Hz. Subjects in these experiments conducted all tasks in one session.

Experiment

Before the start of each trial, the actual target was displayed at the center of the screen for 1 s. After a 1-s delay (blank screen), the search display was shown until the subject found the target or a time-out (20 s) occurred. The subject knew that the target was always present in the search display. This is to exclude possible effects of absence trials. The trial automatically terminated as soon as the subject fixated for at least 320 ms in a radius of 1.5° around the target. Thus, no manual interaction was necessary to terminate the trial. This excludes movement artifacts and speeds-up the search process. Trials were administered in blocks of 24 (Experiment 1) or 36 (Experiment 2) trials. Within a given block, the same two feature dimensions were used.

In any search array, two of the three feature dimensions (color, size, and orientation) were modified whereas the other was kept constant. Thus, there were three different possible tasks: CO, CS, and SO. For any given condition, there are four different search items (e.g., all combinations of red/green and horizontal/vertical for the CO task). Search arrays contained 1 unique target and 48 distractors. The 49 elements were distributed randomly on a 7×7 grid, with 3.25° and 2.25° spacing on the x and y axes, respectively. Noise was added to each grid position (uniformly distributed between $\pm 1^\circ$ and $\pm 0.5^\circ$ on the x and y axes, respectively). There were two types of distractor items, each sharing one feature with the target. Each of the distractor items was present 24 times. Thus, a given search display consisted of three unique items out of the four possible. The left-out item shares no features with the target. Each item occupied between 0.5° and 1.0° (see Figure 1). Items were presented on a light grey background to reduce contrast.

Additionally, we ran several control tasks: three pop-out tasks where the target was unique in one dimension (color,

size, and orientation), one task where only orientation was available (elongated Ts) as a feature, and rotated corners (none of the features available). The elongated Ts task consisted of a long rotated (2°) bar with a small bar at one of its ends (thus looking like elongated Ts). The rotated corners task consisted of corners of equal edge length (four possible corners). We used this task as an upper bound on search performance in the case where none of the three features, color, orientation, and size, was available to guide search. For all control tasks, the target was always shown at the beginning of the trial.

Data analysis

Trials with a time-out (target not found) or eye movements outside the screen were excluded (5.6% of all trials). We analyzed the eye movements in two ways: (i) number of fixations to find the target and (ii) conditional probabilities.

The number of fixations to find the target was used to quantify the degree of difficulty. All fixations made between search screen onset and offset were counted. We find that the average saccade and fixation duration was quite stereotypic for all conditions and subjects (average fixation duration 208 ± 188 ms and average saccade duration 37 ± 14 ms, both $\pm SD$). Thus, the number of fixations to find the target was equal (up to a scale factor) to the time it takes to find the target. This was the case for all subjects except one who was excluded (different fixation durations are dependent on task; thus, this relationship does not hold).

Each fixation was assigned to the item closest to it (distance of fixation location to center-of-mass of the item). We estimated the conditional probability of fixating close to distractors defined by a certain feature, given the target (blue bars in Figures 2B–2D). We calculated these probabilities by counting, for all fixations, how many items (nearest neighbor) shared a certain feature with the target. We repeated the same procedure on data where we randomized the order of trials to estimate chance performance (red bars in Figures 2B–2D). If there was no item within 2° of the fixation, the fixation was classified as “blank” (Figures 2B–2D). We also calculated the conditional probability that the second or the third nearest neighbor of a fixation shares one of the features with the target. For this analysis, the search radius was 5° (instead of 2° , see above).

In some fraction of trials, subjects looked directly at the target but failed to see it; their eyes move away from the target without having rested for the 320 ms required to successfully conclude the trial (see the Results section). We defined trials where this happened as trials where there was a fixation on the target, followed by at least one (often more) fixation away from the target. To exclude small corrective saccades around the target, we excluded trials with less than 4° of the

integrated saccade amplitudes between the last fixation and the on-target fixation.

Computational model

The model (Figure 3) consists of three modules: a first stage that extracts and represents features in the visual scene followed by two parallel stages, one for saccade planning and one for target detection (see sections below). There are six 2-D feature maps corresponding to the two attributes of three feature dimensions each. Each element at location x in the j' feature map, $I^{j'}(x)$, is set to either 0 (absent) or R (present), where $R = 10$ is the baseline rate. The feature maps are combined linearly to yield a rate of a Poisson process at each location:

$$\lambda(x_i) = \sum_{j=1}^6 w_j I^j(x_i). \quad (1)$$

λ can be thought of as a simplified saliency map representation (Itti & Koch, 2001) of the search array, modulated by knowledge about the target. The weights w_j are set based on the knowledge about the target (Wolfe, 1994). By default $w_j = 1$, except for the two feature maps which define the target (red and horizontal in Figure 3). The feature map that represents the *primary* feature (primary with regard to the hierarchy of features that we report; red in the example of Figures 1A and 3) is set to $p + 1$, and the feature map that represents the other feature of the target (*secondary*; here horizontal) is set to $sp + 1$. Thus, $w_1 = p + 1$ and $w_3 = sp + 1$ for the example in Figure 3; p and s are positive parameters. We refer to the weighted sum of all feature maps $\lambda(x)$ as the *target map*. It combines top-down target knowledge with bottom-up visual information and is the basis for all further processing. Each element in the target map is the mean of a Poisson process from which a sample is drawn every time the target map for x is used. In the following, reference to $\lambda(x)$ refers to sampling a number from a Poisson process with mean $\lambda(x)$. Such a random sampling is an important property of any neurobiological circuit and has important consequences.

Computational model—saccade planning

At every fixation, saccade planning decides where to fixate next. This is done by calculating a value $F(x)$ (Equation 2) for every element x and choosing the item x that has the highest value of $F(x)$.

$$F(x) = \lambda(x) + E(r)k_1 - \delta(x, c_t, c_{t-1})k_2. \quad (2)$$

In our search task there were 49 elements; thus, $x = 1 \dots 49$. The saccade planning process has a memory that has

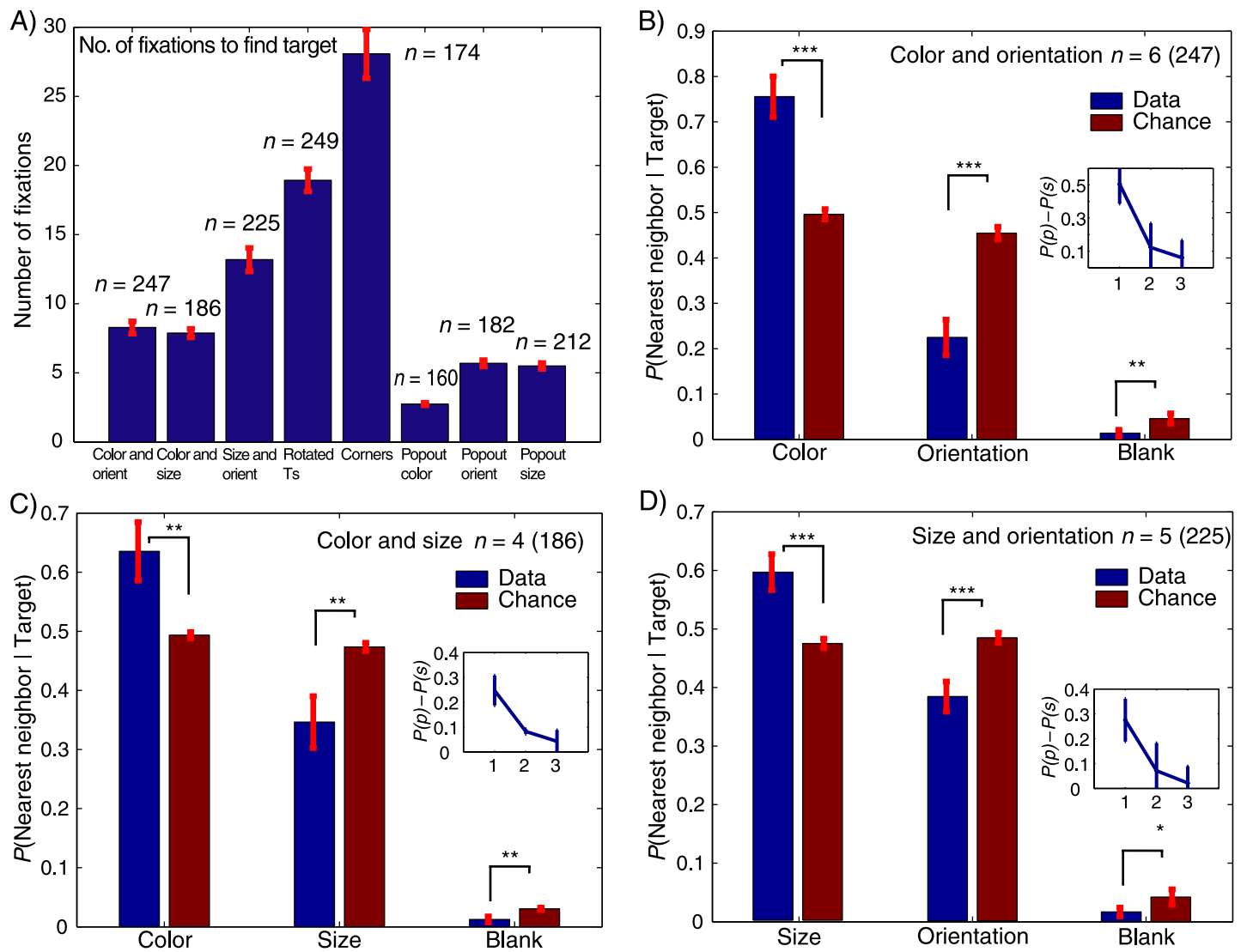


Figure 2. Experimental results and quantification of the search process in terms of difficulty (number of fixations) and conditional probabilities. (A) Number of fixations N to find the target for all conditions. n is the total number of trials. One-way ANOVA with task type (y axis) shows a highly significant effect ($F = 22.43$, $P < 7e - 11$). (B) Conditional probabilities (given the target) of fixating close to a distractor that shares the color or the orientation with the target in the color and orientation (CO) task. A clear color bias is present. Blank corresponds to those fixations when there was no distractor within 2° . (C) Same for the color and size (CS) task. (D) Same for the orientation and size (SO) task. Here, distractors of the same size are preferentially fixated. In panels B–D, n is the number of subjects. Note the high consistency across subjects. The insets in B–D show the difference between the probability of the primary and secondary feature, as a function of the N nearest neighbor. $N = 1$ is the nearest neighbor and is equal to the data shown in the main part of panels B–D. Note that the probabilities are quickly dropping and not significant anymore if the third nearest neighbor is considered (see text for the Discussion section). Error bars in insets are $\pm SD$; all other error bars are $\pm SEM$. ** and *** corresponds to a significance of $<.01$ and $.001$, respectively.

the capacity to store the last m fixated locations (with $m = 0$ corresponding to no memory). If the item with the highest $F(x)$ is currently stored in memory, the next highest value is chosen (iteratively).

The first term in Equation 2 corresponds to the input from the target map and reflects the bottom-up visual signal modulated by target information. The other two terms ensure realistic eye movements. Subjects made saccades that were gamma distributed with a mean of

$5.3^\circ \pm 8.0^\circ$ ($\pm SD$), which is larger than the mean interitem distance of $(2.1^\circ \pm 2.1^\circ)$. To reproduce saccades with such an amplitude distribution, we use a Gaussian “energy” constraint, $E(r) = \exp\left(\frac{-(r - m_{SAD})^2}{k_0}\right)$, with r the distance, in degrees, between the current fixation and the location of x . $E(r)$ is maximal at the preferred saccade length m_{SAD} and is smaller for all other values. m_{SAD} is set equal to the median of the measured distribution of

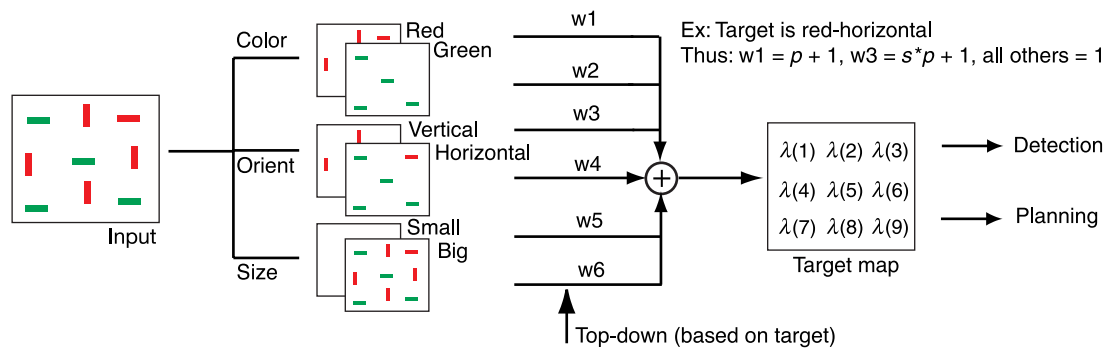


Figure 3. Structure of the model. A feature map is constructed for each of six features (red, green, horizontal, vertical, big, small), containing either a 0 or R for each element. These can be thought of as representing neurons in V4 and in nearby regions of the ventral visual stream. The weighted sum of all feature maps results in the *target map* (right), with one value for every element in the search array. Its value is the mean of a Poisson process from which a number is sampled every time the target map is accessed. The most likely candidate for the target map are neurons in the FEF. Target detection and planning are assumed to be separate processes that receive input from the target map.

the saccade amplitudes (Figure 6C). Furthermore, saccades tend to continue in the same direction as the previous saccade. To reproduce this inertia, we added the term $\delta(x, c_t, c_{t-1})$, corresponding to the angular difference in orientation of two lines: (i) the line connecting the previously fixated location c_{t-1} and the current fixation and (ii) the line between the current fixation location c_t and x . It is normalized, such that the maximal difference (180°) is equal to 1. That is, $\delta(x, c_t, c_{t-1}) = \frac{|\alpha_1 - \alpha_2|}{180}$. $\alpha_1 \leq 180$ and $\alpha_2 \leq 180$ are the orientation of the two lines relative to the horizontal line measured at the origin of the saccades (c_t and c_{t-1}). k_0 , k_1 , and k_2 are constants that set the weight of the two mechanistic terms, relative to the target map.

Computational model—target detection

The target detection process evaluates, at every fixation, whether an item is the target or not. It has the capacity to evaluate C items within D degrees (radius) of the current fixation (both are parameters). If there are more than C items to evaluate within the given area, it evaluates the C items with the highest values of $\lambda(x)$. The capacity limitation does not imply either a parallel or a serial detection process. Whether the detection process has a capacity limitation at all is determined by the value of C : if C is bigger or equal to the number of items within the radius D , the capacity of the process is effectively infinite. We here assume that the target detection process has no memory; that is, it does not take into account any information from previous fixations.

Fit of model to data

We fit the parameters p and s of the model such that both the number of fixations as well as the conditional

probabilities are reproduced as best as possible. The other parameters were kept constant (see the Results section). To find the values of p and s , we calculated the number of fixations $N(p,s)$ and the conditional probabilities $P(p,s)$ for all combinations of p and s between 0.1 and 1 in steps of 0.1. We then used a least squares error measurement to simultaneously fit the two functions $N(p,s)$ and $P(p,s)$.

Results

We first describe the experimental results, followed by the computational model and its application to the experimental data.

Difficulty of search

Figure 1 illustrates typical scanpaths for our display. We quantified task difficulty by how many fixations were required to find the target (Figure 2A). The number of fixations needed to find the target (N) was statistically the same for the two tasks where color was available (CO and CS) (8.19 ± 0.85 and 7.85 ± 0.75 ; $P = .48$, t test), whereas the SO condition required significantly more (13.18 ± 1.72 ; $P < 10^{-6}$, t test). Although this might indicate that orientation cannot be used at all as a feature, additional controls confirm that it can be used if necessary. The two control conditions were rotated Ts (letter T) and rotated corners with equal length of both edges (see Figure 2A and the Methods section). In the “rotated Ts” condition, only orientation was available. Although performance was worse than during a conjunctive search, it was clearly better than in the “corners” condition where search appears to be random (none of the three features available). Also, if any of the three features uniquely

defined the target (e.g., pop-out), the item was quickly found (right three bars in [Figure 2A](#)). This demonstrates that the subjects were able to detect all targets, no matter how they are defined. Note, however, that even in pop-out a substantial number of fixations were required until the target was found (on average 4.45 ± 0.45 fixations, $\pm SE$ with $n = 445$ trials, for all three pop-out tasks).

Guidance of search

Why did certain conditions require more fixations than others (e.g., CS vs. SO)? The items in all three tasks were identical (colored oriented bars). What was different is only which features defined the target. This implies that subjects were able to utilize different features to a different degree. To address this, we estimated the conditional probability of fixating near distractors defined by a certain feature, given the target (blue bars in [Figures 2B–2D](#)). We repeated the same procedure on data where we randomized the order of trials to estimate chance performance (red bars in [Figures 2B–2D](#)). That is, a scanpath associated with one search array for one subject was randomly reassigned to a different search array of the same subject. This procedure was repeated 10 times for each search array. All measured probabilities were different from chance and highly consistent across subjects. For any search condition, the target was defined by two features and there are thus two different conditional probabilities (e.g., $P(\text{share color} \mid \text{Target})$ and $P(\text{share orientation} \mid \text{Target})$ for the CO condition). These two conditional probabilities are not independent because (by design) if an item does not share the color with the target, it shares its orientation and vice versa. We find that subjects primarily used one of the two features in all three search conditions ([Figure 2](#)). If color was available, it was strongly preferred (CO and CS condition, [Figures 2B](#) and [2C](#), respectively); if color defined the target, most eye movements were close to targets whose color was identical to the color of the target. On the other hand, there was a preference for size over orientation in the SO condition ([Figure 2D](#)).

From this, we conclude that there is a strict hierarchy of features: color, size, and orientation. The first feature that defines the target is used to primarily guide search. Thus, color is always used regardless of the other features if it is available. Does this imply that the other features lower in the hierarchy are not used to guide search? This does not seem to be the case because in the two conditions where color was available (CO and CS) the conditional probability for using the other feature (orientation or size) was different (0.24 ± 0.03 vs. 0.37 ± 0.02 ; $P = .01$, t test; [Figures 2B](#) and [2C](#)). If it were ignored, these probabilities should be equal. It thus seems that the primary feature is not the sole factor for determining the fixation probability. Also note that the search difficulty (number of fixations) for the two conditions was not different, despite the different conditional probabilities. On the other hand, the conditional

probability was approximately equal for the CS and SO tasks, but their difficulty was very different (8 vs. 13 fixations). This is a further indication that multiple features are used to guide the search. We will explore this seemingly contradictory fact with our computational model.

The conditional probabilities reported above are calculated based on the item closest to each fixation (the nearest neighbor). How do these probabilities change if they are calculated for the second or third nearest neighbor? The motivation for this analysis stems from the fact that some models propose that saccades are targeted toward groups of items rather than individual items (for example, the center-of-mass). If this were the case, these probabilities should stay similar to the nearest neighbor statistic. We calculated the difference between the conditional probabilities of fixating the primary feature and the secondary feature. This value is high if there is bias toward one feature and approaches zero if there is no bias. Here, we find that this value was positive for nearest neighbors (see above), was barely positive for the second nearest neighbors, and was not different from chance for the third nearest neighbor ([Figures 2B–2D](#), insets; significance for a t test between data and chance control for CO, CS, SO, respectively, for third nearest neighbor: $P = .34, .45, .14$). Thus, only the nearest neighbor seems to induce a strong bias and therefore saccades seem to be primarily targeted toward specific items (see the [Discussion](#) section).

Structure of computational model

We constructed a simple model that reproduces our data and that relates to the known physiological properties of neurons in areas V4 and FEF ([Figure 3](#)). It consists of 2-D feature maps representing the input and the target map, which combines information about the target with the visual input. In the feature extraction stage, an independent map is constructed for every feature ([Figure 3](#), left). These feature maps are combined, biased by information about the target, to create the target map ([Figure 3](#), right). The target map is used to plan where to saccade next and to bias the detection process (for details, see the [Methods](#) section).

The model has five parameters: strength of primary and secondary top-down modulation (p and s), memory capacity (m), detection capacity (C), and detection radius (D). The first two parameters (p and s) modulate the target map. Memory capacity determines how many fixations, relative to the current one, the saccade planning process remembers (and thus does not revisit). The other two parameters influence the detection process (C , D). The saccade planning process is thus parameterized by three parameters only (p , s , and m). Below, we evaluate each parameter in terms of search performance ([Figure 4](#)), quantified by number of fixations (N) as well as conditional probabilities (P). We only plot the conditional probability of the primary feature $P = P(\text{primary} \mid T)$ for

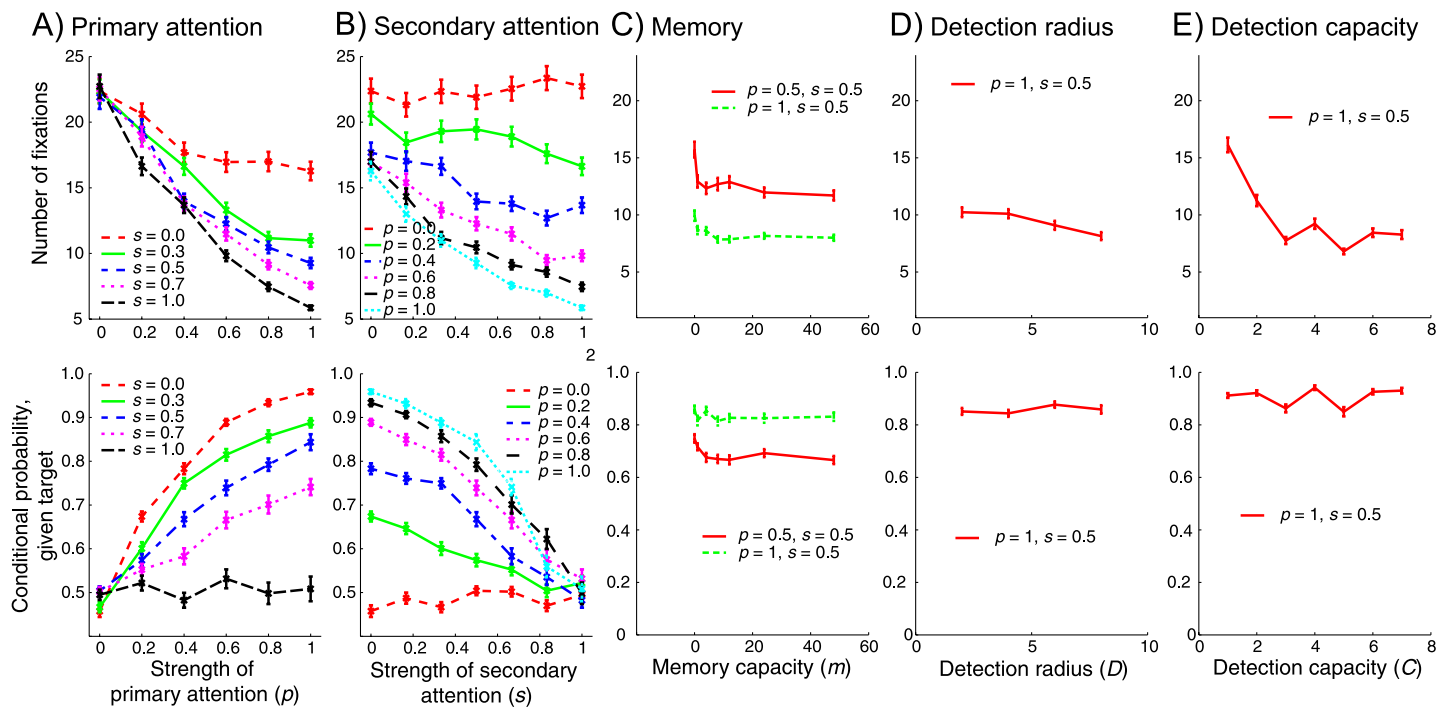


Figure 4. Performance of the model. Only p and s matter. For each of the five model parameters (left to right), two plots are shown (columns): the number of fixations to find the target (N ; top row) and the conditional probability, P , of fixating on the primary feature of the target (bottom row). Only one of the parameters is changed, whereas the others are kept constant. The constant values were $p = 1$, $s = 0.5$, $D = 6$, $C = 2$. See text for more details.

each condition because the conditional probability of the secondary feature is approximately $1 - P$ (except for blank fixations, which are $<1\%$).

Top-down factors determine search performance and strategy

We vary the strength of top-down modulation by the primary and secondary feature by varying p and s each from 0 (no modulation) to 1 (doubling the object's rate in the target map) while keeping all other parameters constant ($m = 1$, $C = 2$, $D = 6$). We specify s as a fraction of p to ensure that primary attention is always stronger than secondary attention; thus, $s = 0.5$ means that s is equal to 50% of p . For example, if $p = 0.6$ and $s = 0.5$, the weights for primary and secondary attention are set to 1.6 and 1.3, increasing the mean amplitude of these two features in the target map by 60% and 30%, respectively.

p and s have a strong, but differential, influence on search performance (Figure 4A). If only one feature is modulated (i.e., $s = 0$), the model cannot be fitted to the data (red line in Figure 4A). For example, for $p = 1$ and $s = 0$, $N > 15$ whereas subjects routinely required less fixations for any of the three tasks (e.g., ~ 8 for the CO and CS task). This is the case despite the very high conditional probability of $p = 1.0$ (higher than measured). Two conclusions can be drawn at this stage: Attentional

modulation to only one feature is not sufficient and strictly fixating on elements that share the primary feature with the target does not guarantee high performance. This also implies that it is necessary to combine the feature maps with a sum instead of a Max operation: The Max method only allows deployment to one feature. In our model, replacing the sum with a Max operation in Equation 1 is equal to setting $s = 0$. As can be seen in Figure 4A, realistic levels of performance are impossible to reach with this setting.

Increasing s while keeping p constant (Figure 4B) increases the number of fixations N and decreases the conditional probability P , reproducing our data. Note that s is a fraction of p and not an absolute value to ensure that $p > s$ at all times (see above). The tradeoff between low and high values of s can be visualized in terms of the target to distractor ratio (TDR) and target visibility. Here, TDR is defined as the ratio between the mean value of the target divided by the mean value of all distractors. The mean value of an item x is equal to the mean value of the Poisson process $\lambda(x)$ in the target map. No secondary attention ($s = 0$) results in maximal difference between the values of the items that share the primary feature with the target and those who do not, but the target is indistinguishable from any other of the 23 distractors of the attended type (TDR = 1). On the other hand, increasing s reduces the difference between the two dimensions of the primary feature map but makes the target more visible

(TDR > 1). This tradeoff explains why lower conditional probabilities for the primary feature do not necessarily result in lower search performance (e.g., CO vs. CS).

Memory capacity of saccade planning only plays a minor role

Varying memory capacity of saccade planning while keeping all other parameters constant (Figure 4C) shows that memory only has a minor influence on search performance. Note that we assume that the detection process has no memory. Changing the value of m only affects the planning process. Increasing the memory capacity from $m = 0$ to $m = 1$ improves performance strongly (Figure 4C). However, further increases in memory capacity ($m > 1$) have a negligible influence on performance. Even perfect memory ($m = 49$) hardly improves performance. See the Discussion section for an explanation of this finding.

Detection capacity and radius

Target detection is executed at each fixation; thus, it only influenced N but not P . First, we explored the influence of C and D on performance under the assumption that the

detection process is capacity limited. Increasing detection capacity while keeping the detection radius constant ($D = 6$) increases performance (Figure 4E), but only until $C = 3$. The detection radius with constant capacity ($C = 2$) only has a minor influence (Figure 4D). Both results can be explained by the properties of the search array: Given even a moderately sized search radius, there are more elements than can be processed within the detection radius. However, only a few (<4) of those candidates are likely to be the target (e.g., share the primary feature) and thus search performance does not increase if C and D are increased further.

Above, we used small values of C and D because we assume that the serial nature of the detection process does not allow more covert attentional shifts within the time of a single fixation. Next, we explored performance in cases of a more capable (possibly parallel, see the Discussion section) detection process ($C > 2, D > 6$). We find (Figure 5) that a higher capacity process leads to large performance improvements if the detection radius is sufficiently large. Realistic detection thresholds are typically below 10° ($D \leq 10$, see the Discussion section). Assuming an infinite detection capacity ($C >$ number of items within radius D), we find that increasing the detection radius from 2 to 10 increases search performance dramatically (Figure 5, $C = 49$). Thus, the model is capable of reproducing a wide range of possible fixation and detection behaviors.

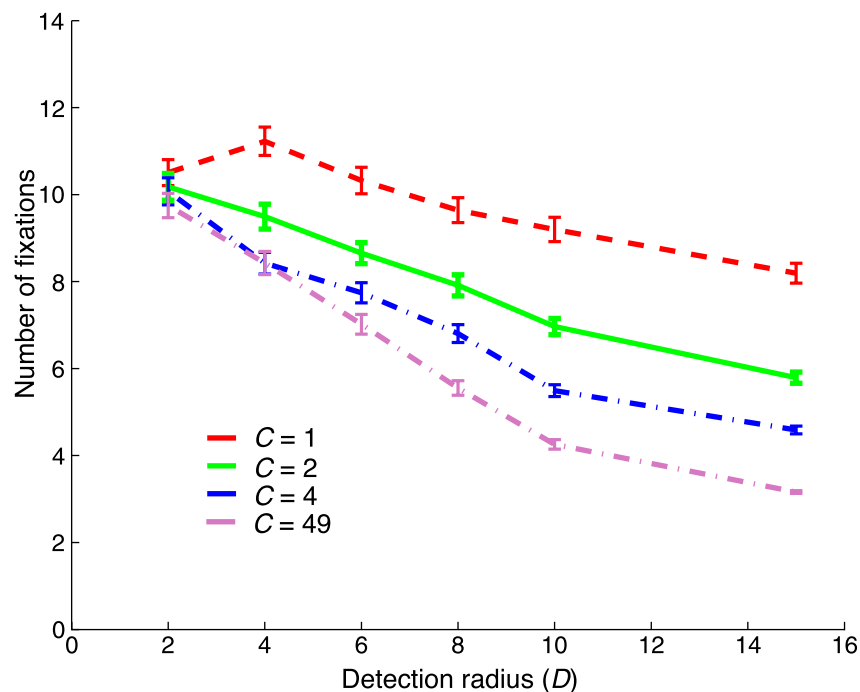


Figure 5. Performance of the model for different parameters of the detection process. Only the number of fixations is plotted because the conditional probabilities are not influenced by the detection process (see Figures 4C and 4D, second row). The green line corresponds to Figure 4D for $C = 2.8$. Note that $C = 49$ is equal to an infinite capacity (= parallel) model. For large values of D , the infinite capacity model has access to all items on the display but still requires a significant number of fixations to find the target. This is due to mechanistic eye movement constraints against unrealistically long saccades.

Fit of model to data

To reproduce our data as shown in Figure 2 (population average of nine subjects), we assume that the last item that was fixated is remembered ($m = 1$), a detection capacity of $C = 2$ and a search radius of $D = 6$. As shown in Figure 4, the exact choice of these parameters is not critical. $k_0 = 15$, $k_1 = 10$, and $k_2 = 5$; these constants are adjusted such that the eye movements have realistic properties (saccade length and angle). Their exact values are not crucial and changing them only affects the mechanistic properties of the generated eye movements.

We reproduced the measured values for N and P for all tasks (CO, CS, SO) by finding values for the parameters p and s that fit the experimental data (see the Methods section). We find that the averaged data can be fit well: Running the model with the identified parameters on the same data that the subjects saw reproduces both the number of fixations as well as the conditional probabilities observed (Figures 6A and 6B). Also, the distribution of the saccade amplitudes is similar to the experimental data (Figure 6C), indicating that the model produces realistic eye movements. Note that the shape of the SAD is not explicitly specified by the model description (Equation 2). Rather, it is a result of the interaction of the term $E(x)$ in Equation 2 and the target map. If evidence for a given item (given by $\lambda(x)$) is strong, long saccades are made. This leads to the long tail of the SAD.

The p/s values for the three tasks (CO, CS, and SO) were 0.9/0.7, 0.9/0.8, and 0.45/0.7. Thus, the increases

in firing rate (mean of Poisson process in model) for the primary and secondary features are 90%/63%, 90%/72%, and 45%/32%, respectively. These values, which were fitted independently for each task, confirm the hierarchy of features that we observed experimentally. Color is the strongest feature (with $P = .9$), followed by size ($P = .45$). Orientation is never the primary feature.

Fixating on the target without seeing it

In some instances (Figure 7A), subjects fixated on the target without seeing it (that is, not stopping the search). To investigate such “return saccades”, we quantified the percentage of trials in which the subjects fixated on the target without stopping the search. We find that, on average, in 12% of all trials of the three tasks (CO, CS, and SO) the target was fixated but the search was not stopped (Figures 7A and 7B). The incidence of such “return to target” trials was higher for the two control conditions and generally seems to be higher for more difficult tasks (Figure 7B). We additionally quantified the number of fixations between an “on-target” fixation and the final target fixation that lead to the trial being successfully concluded (Figure 7B, inset). We find that for most trials, the target was found within a few fixations of the “on-target” fixation (for 77% of on-target trials, three or fewer fixations were between the on-target fixation and the last fixation).

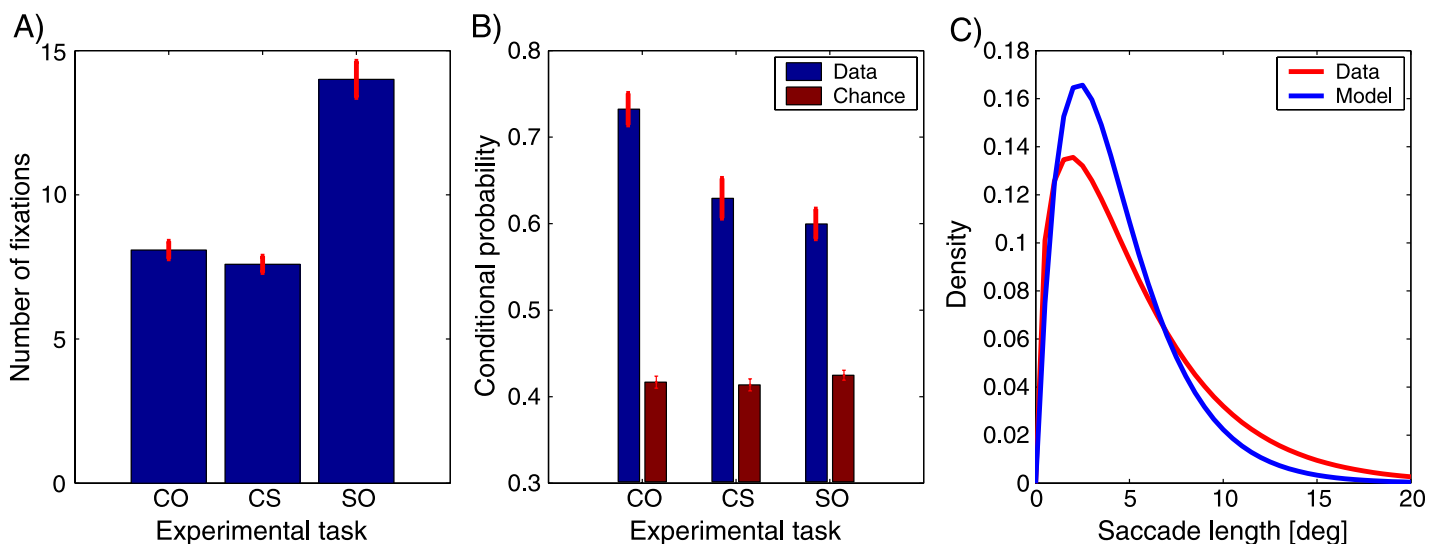


Figure 6. The model fitted against the averaged search performance data of nine subjects. Only p and s were varied whereas all other parameters were kept constant for all simulations (see text). (A) Number of fixations, N . Compare to Figure 2A. (B) Conditional probability, P , of fixating to a distractor sharing the primary feature with the target (here: color, color, size). Compare to Figures 2B–2D. p and s were chosen as following. CO: 0.9, 0.7; CS: 0.9, 0.8; SO: 0.45, 0.7. Thus, the percentage increase in the rate $\lambda(x)$ in the target map was (secondary/primary): CO: 90%/63%; CS: 90%/72%; SO: 45%/32%. (C) Comparison of the SAD distribution of the model and the data.

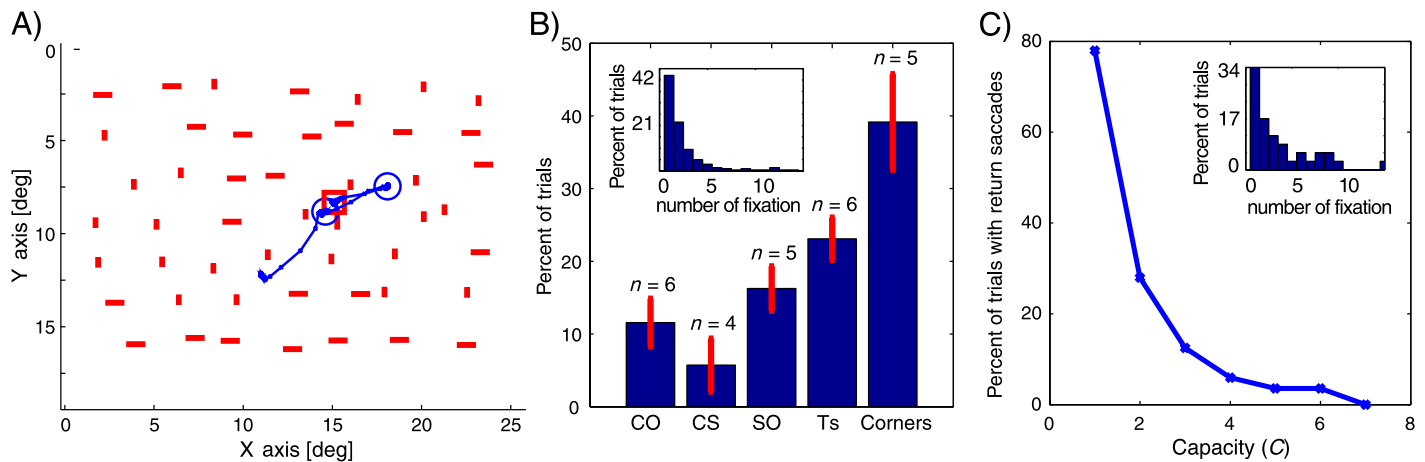


Figure 7. In some trials, the fixation landed on the target, but the subject did not stop the search. Only in a subsequent fixation was the target found. (A) Example of a return to target trial. The first fixation landed on the target, but the subject continued the search. In general, when queried, subjects did not report seeing the target the first time their eyes landed on it. (B) Incidence of fixations on the target in terms of the percentage of all trials (of 1,083 total trials). n are number of subjects. Inset: Distribution of the number of fixations between the on-target fixation and the final target fixation. Seventy-seven percent of all instances had 3 or fewer fixations in between (67% had only 2 or 1). (C) The incidence of return to target trials is faithfully reproduced by the model. Here, the incidence is plotted as a function of the detection capacity. The inset shows the histogram of the number of fixations between the on-target fixation and the final target fixation (as in panel B) at $C = 2$. In most cases (63%), the target is found within three fixations of the on-target fixation. The percentage of trials in the insets in panels B and C are expressed in terms of percentage of all trials with less than 20 fixations between an on-target fixation and the final target fixation.

Fitting the model to individual subjects

What is the variance of the fitted model parameters for individual subjects? In the above analysis, we used the aggregate data from nine subjects to achieve high statistical confidence. Also, we note that the intersubject variability is remarkably small (e.g., error bars in Figure 2 are $\pm SE$ over subjects). In a second experiment, we reran all tasks on three additional subjects on an eye tracker with higher temporal resolution (1,000 Hz, see the Methods section). This enabled us to individually fit the model to each subject and to compare parameters. We find that the individual subjects required a comparable number of fixations for the three tasks CO, CS, and SO. This is illustrated in Figure 8A with error bars as $\pm SEM$. The number of trials for each task that each subject completed successfully varied between 60 and 70. The comparison with the population average (black, from Figure 2A) shows that the individual subjects behaved similarly to the mean of nine subjects (note that the three additional subjects are not part of the population average). One difference, however, is that all our individual subjects were significantly faster for the CS task compared to the CO task (CO vs. CS, $P < .02$ for all 3 subjects, t test). Although the same was the case for the population average, this difference was not significant (number of fixations 8.19 ± 0.85 vs. 7.85 ± 0.75 ; $P = .48$, t test). This indicates that the population average masked an important difference due to intersubject variability. The conditional probability of fixating on the primary feature in each task (color, color,

size for CO, CS, and SO, respectively) was more variable between subjects (Figure 8B). However, the mean values of each conditional probability are well compatible with the population average (black). Next, we will explore differences between individual subjects with our model.

We fit the model parameters p and s (strength of top-down modulation of primary and secondary feature) to each task of each subject (Figures 8C and 8D). All other parameters were kept constant with the values established above: $D = 6$, $C = 2$, and $m = 1$. The mean absolute error between the experimental data and the data produced by the model was small: The difference for the number fixations (N) was 0.16 ± 0.55 fixations ($\pm SD$) and 0.02 ± 0.02 ($\pm SD$) for the probabilities (P). We find that the individual differences between subjects (Figures 8A and 8B) can be accounted for with small variations of the parameters (Figures 8C and 8D). This confirms that individuals are highly consistent for each task, generally having differences in p and s of 10–20%. For example, Subjects 2 and 3 (red, green) have an approximately 55% probability of fixating color instead of size in the CS task (Figure 8B, middle). This is reflected in the parameters in an approximately equal increase for the primary and secondary feature (Figures 8C and 8D), which leads to a lower probability of fixating the primary feature (because both features are modulated equally). However, primary attention is slightly larger in both cases, reflecting the above chance probability of $>.5$ for both. On the other hand, Subject 1 (blue) had a high preference for color in the same task (CS, Figure 8B), which in turn is

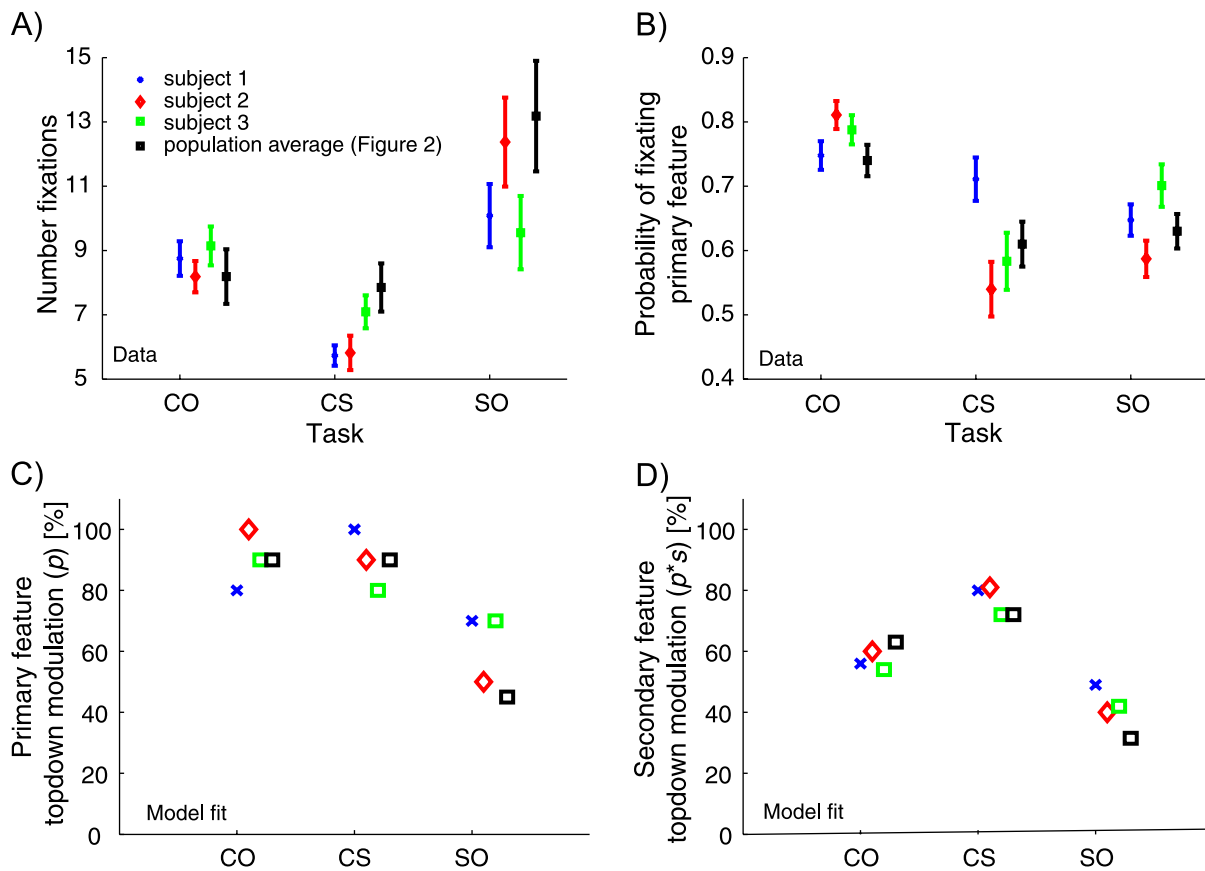


Figure 8. Data and model fits of three individual subjects. (A) Number of fixations for the three tasks CO, CS, and SO for three individual subjects (blue, red, green) and the average population of nine subjects as shown in Figure 2 (black). Note the high consistency between these three subjects. Error bars for individual subjects are $\pm SEM$ with n equal to the number of successfully completed trials for each condition (different for each subject and block, approximately 60–70). Error bars for the population data are $\pm SEM$ with n the number of subjects. (B) Conditional probability of fixating on the primary feature of each of the three tasks (color, color, orientation for CO, CS, and SO, respectively). All probabilities are significantly different from a chance control established by random shuffling (see Figure 2B–2D, red bars). (C and D) Fits of the parameters p and s for each subject (blue, red, green) as well as the population average (black, from Figure 2). Both are specified in terms of absolute increase (%) in firing rate of units representing this feature (here modeled by a mean Poisson rate).

reflected in a higher value of p than s (Figures 8C and 8D; blue in CS). This confirms that our model can be fitted to individual subjects as well as to populations of subjects.

Discussion

With a simple model, we could generate realistic eye movements for a visual search task. The five model parameters correspond to variables which are commonly assumed to be important in visual search (Motter, 1994; Motter & Belky, 1998; Wolfe, 1994). Two had a major influence on search performance: the strength of attentional modulation for the primary and secondary feature (p and s). The strength of attention to the primary feature dominates the number of fixations it takes to find a target, whereas the strength of secondary attention primarily determines the conditional probability of fixation.

Neurophysiological relevance

The structure of the model resembles what we know about saccade generation during visual search. The values in the feature maps can be thought of as representing the mean firing rates of neurons in V4 or nearby regions that are tuned to either color, orientation, or size of the individual bar stimuli. Shortly after stimulus onset, the response of V4 neurons is determined by the visual input, regardless of task relevance. Approximately 150–200 ms afterward, neurons representing relevant stimulus attributes have a higher firing rate (Motter, 1994; Ogawa & Komatsu, 2006). Neurons representing some features of the stimulus (e.g., color) are modulated stronger than others (e.g., orientation) (Bichot et al., 2005; McAdams & Maunsell, 1999a; Motter, 1994). Here, this process is modeled by multiplication with the weights set by p and s . There is good neurophysiological evidence for multiplicative gain control by top-down attentional modulation

of visual cortical neurons (McAdams & Maunsell, 1999a, 1999b; Reynolds, Pasternak, & Desimone, 2000; Treue & Martinez Trujillo, 1999).

Top-down attentional deployment and shifts of gaze are tightly linked (Bichot & Schall, 1999; Corbetta et al., 1998; Moore & Armstrong, 2003; Rizzolatti, Riggio, Dascola, & Umiltà, 1987; Schall & Hanes, 1993). Here, we used eye position (gaze) as an indirect measurement of where attention is currently deployed in space. However, we do not assume that during a fixation attention is restricted to the closest nearby item. In fact, the detection part of our model processes a number of elements (determined by C) in parallel at every fixation, as long as they are within the radius of detection (D parameter). Planning where to fixate next does assume that focal attention shifts to the new location. These assumptions are supported by a number of neurophysiological studies. In particular, neurons located in the FEF are known to be closely related to the initiation of eye movements (Bruce & Goldberg, 1985; Bruce et al., 1985; Schall, Hanes, et al., 1995). The response of FEF neurons is dominated by the visual input that is task relevant, whereas all other input is only weakly represented (regardless of their visual features). The firing rate of an FEF neuron is higher if the item in the RF shares features with the target compared to an item that shares no features with the target (Bichot & Schall, 1999). FEF neurons thus signal, for each item, the estimated probability that this position contains the target. Based on this, it has been proposed that FEF represents an integration of the visual input together with top-down information about the task (Thompson & Bichot, 2005; Thompson, Bichot, & Schall, 2001). Looking at our model, FEF neurons can be thought of as implementing our target map (Figure 3). Thus, each value $\lambda(x)$ in the target map corresponds to the mean firing rate of neurons coding for a particular movement vector (relative to the current fixation). Also, the process of making a saccade where $\lambda(x)$ is maximal has a close neuronal analogy: FEF neurons integrate their input until a threshold is reached (race-to-threshold model; Hanes & Schall, 1996). Thus, the neuron with the largest $\lambda(x)$ will (on average) reach threshold fastest and evoke a saccade.

V4 neurons project directly to FEF and also receive direct feedback from FEF (Schall, Morel, et al., 1995; Stanton et al., 1995). These two areas are tightly linked as demonstrated by microstimulation of neurons in the FEF, which causes changes in the receptive fields of V4 neurons in the same manner as deployment of covert attention to the location spatial location does (Armstrong, Fitzgerald, & Moore, 2006). Thus, neurons in V4 can be thought of as representing the values $\hat{I}(x)$, used in the feature maps of our model.

Also, object selective neurons in IT receive direct input from V4 and are modulated by attention (Chelazzi et al., 1993). These neurons can thus be thought of as representing the detection aspect of the model. Because the input to IT neurons is already attentionally modulated (from V4),

the detection part of the model preferentially evaluates objects which have a high value of $\lambda(x)$.

Psychophysical data also support the notion that the two processes of detection and saccade planning share neuronal pathways: Detection thresholds (measured as d') are the same for both detection (at fixation) and saccades over a wide range of signal-to-noise values as well as tasks (Beutter, Eckstein, & Stone, 2003).

Relationship to visual search

It is well known that knowledge about which features define the target improves search performance and/or accuracy (Burgess & Ghandeharian, 1984; Motter & Belky, 1998; Rajashekar, Bovik, & Cormack, 2006; Rao et al., 2002). Frequently, the time required to find the target is used as a measure of difficulty. This is done by comparing the rate of growth (slope) of the RT as a function of the size of the search array (Treisman, 1988, 1998; Wolfe, 1998). If the time it takes to find a given target is less than to find some other target, it is reasonable to infer that it was easier to search for the target that was found quicker. Because there are many reasons why some targets can be located faster than others, RT measurements alone are not sufficient to constrain mechanistic models. Here, we used a computational model to investigate why it is that some targets are easier to find than others (e.g., CO compared to the SO task). This allowed us to infer that some features (e.g., color) can be used more efficiently for the deployment of top-down attention than others (e.g., orientation). These differences in strength of top-down modulation explained the different amounts of time required to find the target. Thus, recording eye movements during visual search tasks provides additional information that is otherwise not available. Importantly, calculating conditional probabilities (given the target) makes it possible to investigate which feature(s) defining the target were used to guide search to what extent. This is important because this is one of the crucial pieces of information in any model of visual search—given a target, *what* information about the target is used to bias search? Another approach to answer this question is to embed the targets in $1/f$ noise and extract a small patch of the image around each fixation. This patches are then prewhitened (decorrelated) and averaged to yield a classification image (Rajashekar et al., 2006). The classification image shares some (but not all) features with the target. This supports our contention that some features of the target are used preferentially to bias the search.

Models of the detection process

Our detection process is executed at every fixation. There are two principled ways of how the detection process could decide whether the item was present or not: either serially (by deploying covert attention, as in guided

search; GS) or in parallel. Our data do not address this issue and we do not make an assumption as to which process is used. The only assumption is that the items to be processed are within a radius D around the current fixation and that no more than C items can be considered. For large values of C (C larger than the number of items that fit within D), this is equal to a parallel model (Eckstein, Thomas, Palmer, & Shimozaki, 2000; Palmer, 1994; Palmer et al., 2000). For values of C that are smaller than the number of items within the radius D , the model is serial in the sense that it only processes a subset of all possible items. However, we do not make assumptions as to whether the C items processed at every fixation are processed serially (covert attentional shifts) or in parallel. The restricted search radius D is motivated by the fact that the ability to correctly detect the presence or absence of items decreases as a function of distance from the current fixation (Bouma, 1970; Carrasco, Evert, Chang, & Katz, 1995). Depending on the size of the items and signal-to-noise ratio, detection accuracy degrades quickly over a few degrees and is typically smaller than 10° .

Fixating on the target but not seeing it

We found that fixating near the target does not necessarily mean seeing the target (and thus stopping the search). The same phenomena has been observed in monkeys searching for known targets in artificial (Motter & Belky, 1998) and natural scenes (Sheinberg & Logothetis, 2001). This effect can be explained by the capacity limitation of detection. Which items are evaluated is defined by the target map. Because the target map is stochastic, it is possible that the target is not evaluated at fixation even if it is closest to the fixation. One would, however, expect that the target is found within a few fixations of this happening. Indeed, both our subjects as well as the model find the target quickly after the occurrence of such an “on-target” fixation (see the Results section). If the above effect is due to the capacity limitation of the target detection process, one expects the incidence of “return saccades” to decay as a function of the detection capacity of the model (C). Indeed, we find (Figure 7C) that the incidence of such trials is related to C . Assuming that the underlying detection model is a capacity limited model, values of C can be found that produce the same incidence of “return to target” trials as observed experimentally. We used $C = 2$, which somewhat overestimates the incidence of on-target trials.

The incidence of return saccades for $C > 7$ drops to zero (Figure 7C). This occurs because our model does not have an activation threshold (Wolfe, 1994). Thus, arbitrary small activation values in the target map still attract attention. Thus, it is necessary to introduce an activation threshold below which an item is never considered for detection to account for all aspects of the data with a large (possibly infinite) capacity model. The value of this

activation threshold could be fit to the data such that the same “on-target fixation” incidence is reproduced by the model. Our data do not allow us to conclude which is the valid model of detection.

Either model (capacity limited or parallel) is compatible with single cell recordings of object selective cells in inferior-temporal cortex (IT) during such “double-take” trials (Sheinberg & Logothetis, 2001): the IT neuron selective for the search target fails to respond, despite the fact that the monkey is fixating on the object. However, shortly after a saccade is made to another location, the neuron starts responding and the monkey quickly returns to the previously fixated location. The neuron could not have responded either because the serial capacity limited process did not attend to that particular region or because the neurons activity was below the activation threshold.

Bottom-up mechanisms

We did not include lateral interactions (such as center surround inhibition) into our model, preventing us from reproducing the control data that rely on bottom-up effects (e.g., pop-out if a feature is unique). During search (Bacon & Egeth, 1994; Einhauser & Konig, 2003), however, bottom-up effects are generally weak and top-down factors dominate. Furthermore, our model does not take account of interactions among oriented line elements of the sort that will enhance the saliency of contours and other global geometrical arrangement and lead to contour integration and figure-ground segregation (Braun, 1999; Itti, Koch, & Niebur, 1998; Li, 1998; Motter & Holsapple, 2000; Peters, Iyer, Itti, & Koch, 2005).

Comparison with other models

There are many models that attempt to explain how visual search proceeds fixation-by-fixation. We here limit discussion to models that are quantitative and can be used, at least in principle, to generate real eye movement paths. Parts of our model are similar to GS (Wolfe, 1994; Wolfe, Cave, & Franzel, 1989) and related models (Cave, 1999), based on Wolfe’s crucial insight that a bottom-up saliency-like mechanism must be modulated by top-down influences (see, also Bacon & Egeth, 1994; Bacon & Egeth, 1997). Note, however, that GS is a model of covert attention (at fixation). It does not describe eye movements. We model both covert attention (the planning component of our model) as well as overt attention (the detection component). Thus, GS could be used as the detection component of our model (executed at every fixation; see below). In GS, a weight determines how much a particular channel contributes to the activation map; the decision of which weights to change for a particular target are made on grounds of optimality. In contrast, we infer the weights for each feature channel based on the empirically

measured conditional probabilities. There are several crucial differences. (i) Wolfe’s activation map—what we call target map—is calculated once for each search trial (no eye movements). Noise is added and the peaks are rank ordered and visited sequentially (covert) until the target is found or the values are smaller than an activation threshold. Noise is a crucial component of GS: If no noise is added, the target is found immediately on every trial if its activity is higher than the activation threshold (pop-out). In our model, removing noise would eliminate “return saccades” but not lead to pop-out. (ii) GS assumes perfect memory (or inhibition of return) because of the rank ordering (see, however, Horowitz & Wolfe, 1998) whereas memory plays almost no role in our case (see Figure 4C). (iii) Most importantly, our model takes into account several aspects of eye movements. This includes recomputing the target map at every fixation (because the visual input changes) as well as mechanistic constraints of eye movements (typical saccade length and smoothness). (iv) We assume that the information in the target map is represented by a stochastic Poisson process. This has important consequences because different modules (e.g., detection and planning) accessing the same item of the target map do not see the same values. (v) Our model does not include an activation threshold. In contrast to GS, it is thus not capable of generating false alarms nor does it have a systematic way of aborting the search.

Here, a restricted set of features (color, orientation, size) defines each item uniquely. An alternative approach is to use a large set of features defined by a bank of filters (Rao, Zelinsky, Hayhoe, & Ballard, 1996; Rao et al., 2002). In this case, the target is found by fixating the location with the minimal difference in response of the filters to the target and the image patch at each possible location. This process is repeated at each scale (coarse-to-fine) until the target is found. The crucial difference to our approach is that top-down guidance involves every possible feature (fine grained). This also implies perfect memory of the target instruction. In contrast, we only require memory for two features during execution of the task.

The next point that is fixated during the search is the item with the maximal activation in the target map (with some constraints). This approach is in common with most other models of eye movement generation. However, this is not always the case for ideal observer models that take into account the variable d' as a function of eccentricity (Najemnik & Geisler, 2005).

The target map is equal to the weighted linear sum of the feature maps. This approach is commonly used in fixation search models (Baldassi & Verghese, 2002; Eckstein et al., 2000; Palmer et al., 2000; Wolfe, 1994). Here, the weights are provided by what we refer to as top-down attention. A precursor to this approach is dimensional weighting, which suggests that different dimensions can be weighted and combined linearly to produce higher activity for relevant items (Kinchla, Chen, & Evert, 1995; Krummenacher, Muller, & Heller, 2001; Muller et al., 1995;

Murray, Sekuler, & Bennett, 2003). Note, however, that this approach weights dimensions (e.g., “color”) rather than specific features (e.g., “red”). Our data, as well as other psychophysical evidence, suggest that top-down attention is more specific than just the dimension (Navalpakkam & Itti, 2006).

Targeting of saccades

Are fixations attracted by single items or groups thereof? We used the nearest neighbors to estimate the conditional probabilities. However, the conditional probabilities could be biased in two different ways: Either because the fixation was attracted by a single item or because the fixation was attracted to the center-of-mass of a number of items (possibly of the same feature). If fixations would be attracted toward groups of elements sharing the same feature, one would expect that the saccadic bias is (on average) similar for all items which are close to a fixation. However, we found that this is not the case: The conditional probability of the N th nearest neighbor (e.g., $N = 1$ is nearest, $N = 2$ is second nearest) quickly approaches chance for $N > 1$ (Figures 2B–2D, insets). For $N = 3$ the bias is entirely erased and both types of distractors are equally likely to be the third nearest neighbor. Although this does not exclude targeting of saccades toward more complicated ways of grouping, it shows that saccades in our search displays were primarily targeted to single items and not to the center-of-gravity of multiple items. This is particularly relevant in the context of the optimal searcher (Najemnik & Geisler, 2005), which predicts center-of-gravity fixations. However, a suboptimal searcher that always fixates the location with maximum posterior was likewise found to be nearly equal in performance (Najemnik & Geisler, 2005). One extension of GS that takes into account eye movements is the area activation model (Pomplun, Reingold, & Shen, 2003). Here, the activation map is convolved with a 2-D Gaussian to account for the typical “area” that can be processed at every fixation. This model is built on the premise that fixations are guided by groups of items that share a feature with the target (rather than single elements). The number of fixations is equal to the number of peaks in the area activation map. This model assumes that only one feature guides the search (e.g., color). Here, we argue that guidance by both features is necessary to simultaneously account for both the number of fixations and the conditional probability.

An additional factor that influences targeting of saccades is given by the physical limits of the muscles that move the eye. There is a limit to how small or large a saccade can be and typical SADs are distributed very similarly across subjects (Bahill, Adler, & Stark, 1975): the distribution peaks at approximately $2.5\text{--}3^\circ$ and the distribution has a very long tail (up to 20° , see Figure 6C). To include this behavior in our model, we added a

symmetric Gaussian energy term to our model (see Equation 2). This term enforces the preferred saccade length. Its weight, however, is sufficiently low so that it can be overruled by strong items (thus the asymmetry in the resulting SAD, see Figure 6C). Another approach is to predict the SAD entirely from other factors such as limits imposed by acuity (Najemnik & Geisler, 2005). Our model does not attempt such a prediction. Rather, the model assumes fixed physical constraints that partially determine the shape of the SAD.

Comparison with models of visual cortex

Our model is compatible with a feed-forward architecture of visual cortex where higher level processes can modulate gains of certain neurons (Serre, Oliva, & Poggio, 2007). Following the onset of the visual stimulus, V4 and FEF neurons are dominated by the visual input rather than by task relevant information. However, here we have only modeled the steady-state value of V4 and FEF neurons, which is reached after approximately ~ 100 ms in FEF (Thompson, Hanes, Bichot, & Schall, 1996).

Memory

We found that memory beyond the last fixation is not necessary. The memory for the last fixation manifests itself as an inhibition-of-return effect rather than any explicit memory (Klein, 1988; Klein & MacInnes, 1999). Even a perfect memory with $m = 49$ hardly improves performance. This counterintuitive result can be explained by treating search as a random search from an array of n elements with ($m = 0$) or without ($m = n$) replacement. The expected number of draws to find a specific item is n in the former and $n/2$ in the later case. Because the number of fixations to find the target is typically < 10 , the probability that an element that is currently in memory is revisited is small. Thus, memory is not important for this particular search array size and configuration. However, this calculation also indicates that for search arrays with many fewer items that still require a substantial number of fixations, memory might well become important. Note that the constraints of the eye movements (smoothness, typical saccade length) also act as an implicit form of memory. Thus, because saccades tend to follow each other with minimal change in angle (see Figure 1), items previously fixated tend to be avoided even if no explicit memory exists. This additionally decreases the probability of revisiting items without having an explicit form of memory. This might explain some of the previous contradictory results (Horowitz & Wolfe, 1998; Klein, 1988; McCarley et al., 2003). Our model recomputes the entire target map for every fixation and does not have memory for any of the decisions made at the previous fixation. Thus, no trans-saccadic memory integration takes

place, in agreement with a recent Ideal Bayesian Observer model (Najemnik & Geisler, 2005).

Comparison with other eye movement recordings

Many studies recorded eye movements during visual search to assess where fixations preferentially land (Findlay, 1997; Motter & Belky, 1998; Williams & Reingold, 2001; Williams, 1966; Zelinsky, 1996). In agreement with previous studies, we found that color is a feature that strongly guides gaze. One way to quantify selectivity of fixations is to calculate the frequency by which each type of distractor attracts fixations. Here, we quantify the same effect by calculating the conditional probability that, given the target, a certain feature will be shared by the nearest neighbors of the fixations. Fixation frequency and conditional probability are different in that we allow the probabilities to change as a function of the target (conditional). Indeed, we find that the probability of fixating the very same distractor elements varies greatly as a function of the target instruction. There has been some controversy as to how strong certain features guide attention: Most studies found that color is a strongly guiding feature (Motter & Belky, 1998; Williams & Reingold, 2001; Williams, 1966), whereas at least one report the opposite (Zelinsky, 1996). Here, we find that such differences can occur as a result of which features define the target. In fact, in our model, weak preference for one feature versus another does not necessarily imply weak top-down attentional priors. Using the computational model, we showed that this is because of strong top-down attentional biases to multiple features, which results in weak selectivity and high performance at the same time.

Prediction of gain modulation from eye movement data

Our model predicts the strength of gain modulation of V4 neurons during visual search; modulation will be strongest for color and smallest for orientation, with intermediate values for size. Using items defined by color and luminance (Motter, 1994) found that the firing rate of V4 neurons was approximately double when the stimulus in the receptive field matched the target compared to when it did not match the target. This corresponds to $p = 1$ in our model. Using more complex items (cartoons of objects) (Chelazzi, Miller, Duncan, & Desimone, 2001) found a modulation by the search target of between 39% and 63% (in terms of our model, this corresponds to a variation of P of .39–.63). Others (Bichot et al., 2005) discovered that the modulation for V4 neurons responsive to color was much stronger than modulation of neurons responsive to shape (e.g., mostly orientation). Another study found that V4 neurons tuned to orientation increased

firing by on average only 20% (McAdams & Maunsell, 1999a), which is much less compared to color. Although it seems puzzling that some features are modulated less than others, our model predicts exactly this situation. We also show that this is the optimal strategy to use: Increasing firing for one feature more than the other results in better performance than strongly modulating only one feature. It remains an open question why color rather than, say, orientation has primacy in terms of strength of top-down modulation.

Conclusions

Measuring human eye movements is considerably easier and faster than recording neuronal responses in monkeys. However, models of attention depend on how neurons in areas such as V4 or FEF are modulated. Here, we use a simple model that can be fitted to measured eye movement data to deduct the gain modulation of V4 neurons. This model can be used to predict the modulation of firing rates of V4 (and FEF) neurons for arbitrary tasks where the target and distractor items are defined by a well-restricted set of features.

Acknowledgments

We would like to acknowledge the insightful comments of the anonymous reviewers that greatly benefited the paper. We would like to thank Ralph Adolphs for providing the Eyetracker equipment we used in the first experiment and Dirk Neumann and Wolfgang Einhaeuser for discussion. This work was supported by the National Geospatial Intelligence Agency (NGA), NIMH, NSF, ONR, and DARPA.

Commercial relationships: none.

Corresponding author: Ueli Rutishauser.

Email: urut@klab.caltech.edu.

Address: California Institute of Technology, MC 136-93, Pasadena, CA 91125, USA.

References

- Armstrong, K. M., Fitzgerald, J. K., & Moore, T. (2006). Changes in visual receptive fields with microstimulation of frontal cortex. *Neuron*, *50*, 791–798. [[PubMed](#)]
- Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception & Psychophysics*, *55*, 485–496. [[PubMed](#)]
- Bacon, W. J., & Egeth, H. E. (1997). Goal-directed guidance of attention: Evidence from conjunctive

visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 948–961. [[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](#)]
- Baldassi, S., & Verghese, P. (2002). Comparing integration rules in visual search. *Journal of Vision*, *2*(8):3, 559–570, <http://journalofvision.org/2/8/3/>, doi: 10.1167/2.8.3. [[PubMed](#)] [[Article](#)]
- Beutter, B. R., Eckstein, M. P., & Stone, L. S. (2003). Saccadic and perceptual performance in visual search tasks: I. Contrast detection and discrimination. *Journal of the Optical Society of America A, Optics, image science, and vision*, *20*, 1341–1355. [[PubMed](#)]
- Bichot, N. P., Rossi, A. F., & Desimone, R. (2005). Parallel and serial neural mechanisms for visual search in macaque area V4. *Science*, *308*, 529–534. [[PubMed](#)]
- Bichot, N. P., & Schall, J. D. (1999). Effects of similarity and history on neural mechanisms of visual selection. *Nature Neuroscience*, *2*, 549–554. [[PubMed](#)] [[Article](#)]
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, *226*, 177–178. [[PubMed](#)]
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. [[PubMed](#)]
- Braun, J. (1999). On the detection of salient contours. *Spatial Vision*, *12*, 211–225. [[PubMed](#)]
- Bruce, C. J., & Goldberg, M. E. (1985). Primate frontal eye fields: I. Single neurons discharging before saccades. *Journal of Neurophysiology*, *53*, 603–635. [[PubMed](#)]
- Bruce, C. J., Goldberg, M. E., Bushnell, M. C., & Stanton, G. B. (1985). Primate frontal eye fields: II. Physiological and anatomical correlates of electrically evoked eye movements. *Journal of Neurophysiology*, *54*, 714–734. [[PubMed](#)]
- Burgess, A., & Ghandeharian, H. (1984). Visual signal detection. I. Ability to use phase information. *Journal of the Optical Society of America A, Optics and image science*, *1*, 900–905. [[PubMed](#)]
- Carrasco, M., Evert, D. L., Chang, I., & Katz, S. M. (1995). The eccentricity effect—Target eccentricity affects performance on conjunction searches. *Perception & Psychophysics*, *57*, 1241–1261. [[PubMed](#)]
- Cave, K. R. (1999). The FeatureGate model of visual selection. *Psychological Research*, *62*, 182–194. [[PubMed](#)]
- Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature*, *363*, 345–347. [[PubMed](#)]
- Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (2001). Responses of neurons in macaque area V4

- during memory-guided visual search. *Cerebral Cortex*, *11*, 761–772. [PubMed] [Article]
- Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., et al. (1998). A common network of functional areas for attention and eye movements. *Neuron*, *21*, 761–773. [PubMed] [Article]
- Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers*, *34*, 613–617. [PubMed]
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222. [PubMed]
- Eckstein, M. P., Thomas, J. P., Palmer, J., & Shimozaki, S. S. (2000). A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Perception & Psychophysics*, *62*, 425–451. [PubMed]
- Einhauser, W., & Konig, P. (2003). Does luminance-contrast contribute to a saliency map for overt visual attention? *European Journal of Neuroscience*, *17*, 1089–1097. [PubMed]
- Findlay, J. M. (1997). Saccade target selection during visual search. *Vision Research*, *37*, 617–631. [PubMed]
- Gallant, J. L., Connor, C. E., Rakshit, S., Lewis, J. W., & Van Essen, D. C. (1996). Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *Journal of Neurophysiology*, *76*, 2718–2739. [PubMed]
- Hanes, D. P., & Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science*, *274*, 427–430. [PubMed]
- Horowitz, T. S., & Wolfe, J. M. (1998). Visual search has no memory. *Nature*, *394*, 575–577. [PubMed]
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*, 194–203. [PubMed]
- 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.
- Kinchla, R. A., Chen, Z., & Evert, D. (1995). Precue effects in visual search: data or resource limited? *Perception & Psychophysics*, *57*, 441–450. [PubMed]
- Klein, R. (1988). Inhibitory tagging system facilitates visual search. *Nature*, *334*, 430–431. [PubMed]
- Klein, R. M., & MacInnes, W. J. (1999). Inhibition of return is a foraging facilitator in visual search. *Psychological Science*, *10*, 346–352.
- Krummenacher, J., Muller, H. J., & Heller, D. (2001). Visual search for dimensionally redundant pop-out targets: Evidence for parallel-coactive processing of dimensions. *Perception & Psychophysics*, *63*, 901–917. [PubMed]
- Li, Z. (1998). A neural model of contour integration in the primary visual cortex. *Neural Computation*, *10*, 903–940. [PubMed]
- McAdams, C. J., & Maunsell, J. H. (1999a). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *Journal of Neuroscience*, *19*, 431–441. [PubMed] [Article]
- McAdams, C. J., & Maunsell, J. H. (1999b). Effects of attention on the reliability of individual neurons in monkey visual cortex. *Neuron*, *23*, 765–773. [PubMed] [Article]
- McCarley, J. S., Wang, R. F., Kramer, A. F., Irwin, D. E., & Peterson, M. S. (2003). How much memory does oculomotor search have? *Psychological Science*, *14*, 422–426. [PubMed]
- Moore, T., & Armstrong, K. M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature*, *421*, 370–373. [PubMed]
- Moore, T., & Fallah, M. (2004). Microstimulation of the frontal eye field and its effects on covert spatial attention. *Journal of Neurophysiology*, *91*, 152–162. [PubMed] [Article]
- Motter, B. C. (1994). Neural correlates of attentive selection for color or luminance in extrastriate area V4. *Journal of Neuroscience*, *14*, 2178–2189. [PubMed] [Article]
- Motter, B. C., & Belky, E. J. (1998). The guidance of eye movements during active visual search. *Vision Research*, *38*, 1805–1815. [PubMed]
- Motter, B. C., & Holsapple, J. W. (2000). Cortical image density determines the probability of target discovery during active search. *Vision Research*, *40*, 1311–1322. [PubMed]
- Muller, H. J., Heller, D., & Ziegler, J. (1995). Visual-search for singleton feature targets within and across feature dimensions. *Perception & Psychophysics*, *57*, 1–17. [PubMed]
- Murray, R. F., Sekuler, A. B., & Bennett, P. J. (2003). A linear cue combination framework for understanding selective attention. *Journal of Vision*, *3*(2):2, 116–145, <http://journalofvision.org/3/2/2/>, doi:10.1167/3.2.2. [PubMed] [Article]
- Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, *434*, 387–391. [PubMed]
- Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, *45*, 205–231. [PubMed]

- Navalpakkam, V., & Itti, L. (2006). Top-down attention selection is fine grained. *Journal of Vision*, 6(11):4, 1180–1193, <http://journalofvision.org/6/11/4/>, doi:10.1167/6.11.4. [PubMed] [Article]
- Ogawa, T., & Komatsu, H. (2006). Neuronal dynamics of bottom-up and top-down processes in area V4 of macaque monkeys performing a visual search. *Experimental Brain Research*, 173, 1–13. [PubMed]
- Palmer, J. (1994). Set-size effects in visual search: The effect of attention is independent of the stimulus for simple tasks. *Vision Research*, 34, 1703–1721. [PubMed]
- Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research*, 40, 1227–1268. [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]
- Pomplun, M., Reingold, E. M., & Shen, J. Y. (2003). Area activation: A computational model of saccadic selectivity in visual search. *Cognitive Science*, 27, 299–312.
- 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://journalofvision.org/6/4/7/>, doi:10.1167/6.4.7. [PubMed] [Article]
- Rao, R. P., & Ballard, D. H. (1997). Dynamic model of visual recognition predicts neural response properties in the visual cortex. *Neural Computation*, 9, 721–763. [PubMed]
- Rao, R. P., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (1996). Modeling saccadic targeting in visual search. In D. Touretzky, M. Mozer, & M. Hasselmo (Eds.), *Advances in Neural Information Processing Systems* (vol. 8, pp. 830–836). Cambridge, MA: MIT Press.
- 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]
- Reynolds, J. H., Pasternak, T., & Desimone, R. (2000). Attention increases sensitivity of V4 neurons. *Neuron*, 26, 703–714. [PubMed] [Article]
- Rizzolatti, G., Riggio, L., Dascola, I., & Umiltà, C. (1987). Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia*, 25, 31–40. [PubMed]
- Schall, J. D., & Hanes, D. P. (1993). Neural basis of saccade target selection in frontal eye field during visual search. *Nature*, 366, 467–469. [PubMed]
- Schall, J. D., Hanes, D. P., Thompson, K. G., & King, D. J. (1995). Saccade target selection in frontal eye field of macaque: I. Visual and premovement activation. *Journal of Neuroscience*, 15, 6905–6918. [PubMed] [Article]
- Schall, J. D., Morel, A., King, D. J., & Bullier, J. (1995). Topography of visual cortex connections with frontal eye field in macaque: Convergence and segregation of processing streams. *Journal of Neuroscience*, 15, 4464–4487. [PubMed] [Article]
- Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 6424–6429. [PubMed] [Article]
- Sheinberg, D. L., & Logothetis, N. K. (2001). Noticing familiar objects in real world scenes: The role of temporal cortical neurons in natural vision. *Journal of Neuroscience*, 21, 1340–1350. [PubMed] [Article]
- Stanton, G. B., Bruce, C. J., & Goldberg, M. E. (1995). Topography of projections to posterior cortical areas from the macaque frontal eye fields. *Journal of Comparative Neurology*, 353, 291–305. [PubMed]
- Swensson, R. G., & Judy, P. F. (1981). Detection of noisy visual targets—Models for the effects of spatial uncertainty and signal-to-noise ratio. *Perception & Psychophysics*, 29, 521–534. [PubMed]
- Thompson, K. G., & Bichot, N. P. (2005). A visual salience map in the primate frontal eye field. *Progress in Brain Research*, 147, 251–262. [PubMed]
- Thompson, K. G., Bichot, N. P., & Schall, J. D. (2001). From attention to action in frontal cortex. In J. Braun, C. Koch, & J. Davis (Eds.), *Visual attention and cortical circuits* (pp. 137–157). Cambridge: MIT Press.
- Thompson, K. G., Hanes, D. P., Bichot, N. P., & Schall, J. D. (1996). Perceptual and motor processing stages identified in the activity of macaque frontal eye field neurons during visual search. *Journal of Neurophysiology*, 76, 4040–4055. [PubMed]
- Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology A*, 40, 201–237. [PubMed]
- Treisman, A. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, 353, 1295–1306. [PubMed] [Article]
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97–136. [PubMed]
- Treue, S., & Martinez Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, 399, 575–579. [PubMed]
- Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y. H., Davis, N., & Nuflo, F. (1995). Modeling visual-

- attention via selective tuning. *Artificial Intelligence*, 78, 507–545.
- Verghese, P. (2001). Visual search and attention: A signal detection theory approach. *Neuron*, 31, 523–535. [[PubMed](#)] [[Article](#)]
- Williams, D. E., & Reingold, E. M. (2001). Preattentive guidance of eye movements during triple conjunction search tasks: The effects of feature discriminability and saccadic amplitude. *Psychonomic Bulletin & Review*, 8, 476–488. [[PubMed](#)]
- Williams, L. G. (1966). Effect of target specification on objects fixated during visual search. *Perception & Psychophysics*, 1, 315–318.
- Wolfe, J. M. (1994). Guided Search 2.0—A revised model of visual-search. *Psychonomic Bulletin & Review*, 1, 202–238.
- Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science*, 9, 33–39.
- Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 419–433. [[PubMed](#)]
- Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, 5, 495–501. [[PubMed](#)]
- Zelinsky, G. J. (1996). Using eye saccades to assess the selectivity of search movements. *Vision Research*, 36, 2177–2187. [[PubMed](#)]