Human Search for a Target on a Textured Background is Consistent with a Stochastic Model

A. D. F. Clarke[a], P. R. Green[b], M. J. Chantler[b], A. R. Hunt[a]

[a]School of Psychology, King's College, University of Aberdeen, Aberdeen, United Kingdom
[b]School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, United Kingdom

Abstract

Here we compare results from both a stochastic and an optimal model to human behaviour in a challenging visual search task. Both the models rely on a target visibility map based on human performance in a separate detection task. The stochastic model randomly selects the next saccade in the sequence from empirical data, while the optimal model fixates the location that will maximise the probability of detecting the target. As with previous work, we find that the number of fixations required by human observers to locate a target on a textured background is consistent with predictions from an optimal model. However, we find that a memoryless stochastic model also matches human performance. The distributions of saccade amplitudes and directions produced by the stochastic model are more similar to human behaviour than the optimal model. We conclude that, when searching for a target in noise, humans use an essentially random strategy, which achieves near optimal behaviour due to biases in the distributions of saccades we have a tendency to make. The findings reconcile the existence of highly efficient search human performance with recent studies demonstrating clear failures of optimality in single and multiple saccade tasks.

1. Introduction

The human retina provides highly accurate and detailed central vision, but acuity diminishes rapidly with eccentricity. Eye movements shift new locations to central vision, and in doing so sequentially sample finer-grained details from locations that are likely to yield important information, presumably using some combination of peripheral visual signals, inferences based on context, and top-down strategies. Each eye movement during extended search can therefore be useful for understanding how the visual system combines and prioritises information both within each fixation and across a sequence of fixations.

Much of the research on visual search to date has formalised this general issue by focusing on questions of *feature extraction* and of *strategy*. Feature ex-

---

traction includes both top-down guided search (Wolfe, 2007; Zelinsky, 2008) and stimulus-driven (saliency) effects (Itti and Koch, 2000; Gao, Mahadevan, and Vasconcelos, 2008; Itti and Baldi, 2009). For the abstract and discrete search items commonly used as visual search stimuli, categorical features such as colour, orientation, shape and size are often used. Simple qualitative comparisons between the search items and the target can be used to model top-down guidance (Pomplun, Shen, and Reingold, 2003; Rutishauser and Koch, 2007). For more complex stimuli, such as a target hidden in image noise or in a photograph of a natural scene, there is no discrete set of items to consider and more sophisticated image processing techniques are required (Rao, Zelinsky, Hayhoe, and Ballard, 2002; Zelinsky, 2008; Pomplun, 2007; Hwang, Higgins, and Pomplun, 2009; Tavassoli, van der Linde, and Bovik, 2009). In either case, the output of a feature extraction mechanism is an activation map; that is, a representation of the visual array in which peaks of activity represent the priority of locations for attention.

Strategy refers to the mechanism for selecting which location to inspect next. While a number of different mechanisms have been put forward, the most commonly implemented has been the Maximum A Priori (MAP) observer. The MAP observer directs saccades to the current maximum of the activation map and a simple inhibition of return (IOR) mechanism is used to stop the model returning to previously fixated maxima. Depending on the model, a maximum will either represent a search item, or the centre of gravity of a number of search items. As most previous computational models have primarily been interested in the feature extraction stage of search, the MAP observer has often been used for simplicity (Itti and Koch, 2000; Rao et al., 2002; Pomplun et al., 2003; Rutishauser and Koch, 2007; Clarke, Green, and Chantler, 2009; Zelinsky, 2008).

An alternative to the MAP Observer is the Ideal Observer. Here, eye movements are directed to locations that are likely to yield the most information. An example of an Ideal Observer model comes from Najemnik and Geisler (2005), who measured visual sensitivity to a Gabor patch in varying amounts of noise across a range of eccentricities and angles from fixation. From the visual sensitivity data they could generate a model of optimal eye movement behaviour that selected as the next fixation the location that would maximize the probability of detecting the target, given the amount of background noise and the known visibility of the target at various eccentricities. The number of fixations made during search for the target by the human observers (the two authors) closely matched the optimal model. In a second study (Najemnik and Geisler, 2008) they also measured the fixations generated by an optimal model and found that, when averaged over all trials, the Ideal Observer matched the human spatial distribution of fixations: both the model and human observers exhibited a preference for fixating above and below the centre of the image. The idea that eye movements during search are near-optimal is broadly consistent with studies demonstrating the speed and efficiency with which eye movements can be directed to locations in a naturalistic setting that provide the most task-relevant information. A now-classic example is the demonstration of expert cricket batsmen's ability to shift their eyes rapidly to the anticipated bounce

point of the ball based on its trajectory as it leaves the bowler's hand (Land and McLeod, 2000). This is a specific example of a number of studies demonstrating that eye movements are tightly constrained by task goals, and driven to maximise task-relevant information gain (for a recent review, see Hayhoe and Ballard (2014)).

In contrast with the notion that humans are close to optimal in search behaviour is recent evidence of sub-optimality in a very similar context. Morvan and Maloney (2012) instructed observers to first make a single eye movement to their choice of one of three squares aligned in a row, and to then make a judgement about a dot that could appear in either the leftmost or the rightmost square. When the squares are closely spaced, the centre location is the optimal choice because the dot will be visible whether it appears in the left or the right location. As the distance between the squares increases, a point is reached where the centre location is no longer optimal; instead observers can maximize accuracy by selecting either the left or right location. A single saccade in this experiment represents a very similar decision to each saccade in the search task of Najemnik and Geisler (2005), in that observers must use knowledge about their own visual acuity to guide their eyes to the location that is likely to yield the most information. Nonetheless, observers in Morven and Maloney's experiment were far from optimal: not only did observers not change strategy at an optimal spacing, they did not adapt their strategy to changes in the spacing of the squares at all, even though observers were given a monetary reward for each correct response. A similar conclusion was reached by Verghese (2012), who demonstrated that observers failed to adapt their visual search strategies to take target probability information into account, and by Zhang, Morvan, Etezad-Heydari, and Maloney (2012) who found sub-optimal eye-hand coordination in a reaching task.

How can these demonstrations of sub-optimal eye movement behaviour be reconciled with Najemnik and Geisler (2005, 2008)? It is unlikely that observers would be sub-optimal at the level of a single saccade but optimal across multiple saccades. Morvan and Maloney (2012) suggest that observers may adopt heuristics during search that generate sequences of saccades that appear optimal, but are not actually based on a fixation-by-fixation computation of posterior probability of target location. General tendencies observed in search scan-paths that have been taken to be indicative of optimal behaviour may instead be biases in saccade selection related to the scene statistics, the location of the eyes within the scene boundaries, and local mechanisms like inhibition of return and saccadic momentum. For example, Over, Hooge, Vlaskamp, and Erkelens (2007) have shown that search scan-paths exhibit coarse-to-fine structures, that is, observers make shorter saccades as search progresses. Over, Hooge, and Erkelens (2003) found that saccade directions are influenced by the edges of the search image and reported a preference for making saccades parallel to the boundaries of the stimuli. Gilchrist and Harvey (2006) argue that the presence of a horizontal bias in saccade directions indicates systematic scanning in visual search. They suggest these systematic tendencies can be hard to detect in scan-paths because of interactions with salience-based object selection.

3

Chance has also been demonstrated to play a significant role in visual search performance. Using saccade amplitude distributions Motter and Holsapple (2001) calculated the probability of fixating the target by chance under different conditions. While this chance component decreases as the number of distracters increase, it continues to account for a sizeable fraction of performance. Random walks have been successfully used to model an observer's speed and accuracy in present/absent forced choice experiments (Stone, 1960; Reeves, Santhi, and DeCaro, 2005). Rather than model the spatial distribution of fixations these models simulate the observer's decision making process. The random walk occurs between two boundaries (one for a target present response and one for target absent), and is governed by a drift and bias.

A plausible alternative to the optimal model of human search behaviour is therefore that natural search behaviour is stochastic, but constrained by both scene statistics and heuristics. Here we directly compare the performance of a random-walk model to human eye movements during search of a textured surface for an indentation (Clarke, Green, Chantler, and Emrith (2008), see Figure 1 for an example). We used textured surfaces because they appear naturalistic but, unlike photographs of natural scenes, they are fully controlled and parametrised. Our stochastic model randomly selects the next saccade in the sequence from the total set of saccades made from that region of the search array. This model captures the global biases of saccade programming during search reviewed above, but, unlike the optimal model, it does not take into account previous fixations or the visibility of the target given the roughness of the surface texture. For comparison, we also implemented an optimal model, based loosely on that used by Najemnik and Geisler (2005). Each fixation during search was determined based on maximising the amount of information gained about the probable target location. Information gain was dependent on visual sensitivity at different eccentricities and different amounts of surface roughness, established using a separate target detection experiment. The results demonstrate that the stochastic model and the optimal model both closely match the number of fixations required to detect the target in human data. However, the scan paths produced by the stochastic model are far more similar to human search than the optimal model.

2. Experiment 1

In order to compare the search performance of human observers to a random walk, we carried out an experiment with a group of nine observers, all naive to the aim of the study. There are two main parts to the study (*target detection* and *visual search*).

The goal of the target detection part of the study was to generate a target visibility map to be used in both the stochastic and the optimal search models to determine for each fixation whether or not the target has been found. In order to match as closely as possible the detection task observers would need to perform during each fixation while searching for the target, we presented the target in the target detection task at any one of 64 locations. This is

a departure from the method used by Najemnik and Geisler (2005, 2008), in which the target visibility maps were generated from a target detection task in which the target location was cued on every trial and then presented on half of the trials. Certainty about the target location allows covert attention to be allocated to the region of the target, which has been demonstrated to increase visual sensitivity (Yeshurun and Carrasco, 1998).

Although we believe an uncertain target location provides a better estimate for the visibility of the target during search, the downside of our method is that false positives are problematic. While in Najemnik and Geisler (2005, 2008) a false positive rate could be calculated for a given eccentricity, in our method a "target absent" response cannot be linked to a particular location. To cope with this, our experiment was designed to minimise the false positive rate. Participants were aware that the target was present on nearly every trial, and were instructed not to guess, but to respond whenever they saw the target. The experiment included catch trials with no target presented on which feedback was provided to discourage guessing. Two participants were not included in the study because their false positive rates exceeded 15% in the target detection session.

In the visual search part of the study, a similar strategy was encouraged in our observers: the target was present on nearly every trial, and observers were encouraged to search until they found the target. Catch trials with feedback were included to discourage guessing.

2.1. Methods

2.1.1. Observers

Nine 20-29 year old observers (mean age 23.8 years) with corrected/corrected-to-normal vision took part in the experiment. All were naive to the purpose of the study. Two of the participants were undergraduate RAs and completed the Visual Search session first, and the Target Detection session second. The remaining eight observers were paid £5 for the visual search experiment (approximately 45 minutes), and £15 for the target detection part (1hr30-2hr30). Four of these participants carried out the Visual Search experiment first, while the other four carried out the Target Detection session first. All gave informed consent to participate in the experiment, which was approved by the Aberdeen School of Psychology ethics committee.

2.1.2. Surface Stimuli and Equipment

A range of rough surfaces were generated by applying Lambert's cosine law to height maps generated by a $1/f^{\beta}$-noise process. For full technical details see Clarke et al. (2008). The surface roughness is governed by β and a scaling factor, RMS roughness, which was kept constant, $\sigma_{RMS} = 1.1$. The three levels of surface roughness created by varying β will be referred to as `smooth` ($\beta = 1.70$), `medium` ($\beta = 1.65$) and `rough` ($\beta = 1.60$). By changing the random seed used to create the noise we can create textured surfaces on each trial that are unique but statistically identical. The target was created by subtracting an ellipsoid from the three dimensional surface. Examples are shown in Figure 1.
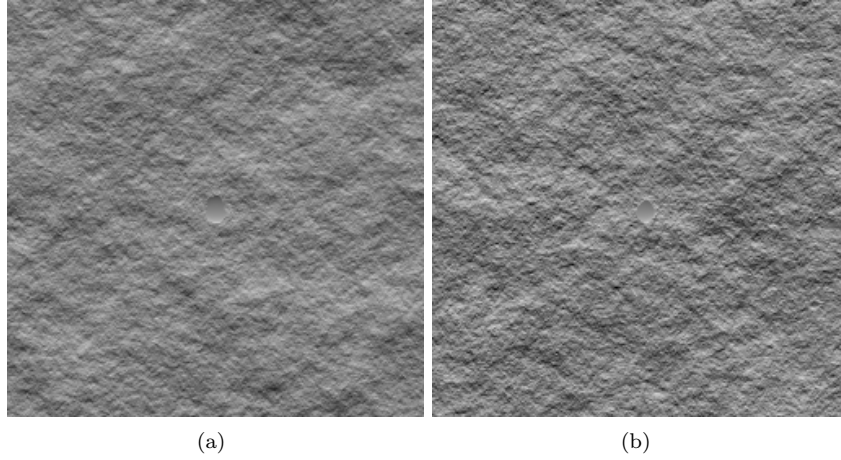
5

Figure 1: Example stimuli. This is a 256×256 pixel crop of one of the (a) smooth and (b) rough surfaces. In both cases the target is shown in the centre of the image. The stimuli used in the experiment were 1024×1024 pixels in size, making the target much smaller relative to the search area than is shown here. The slight differences in the target's shape is due to randomness in the surface at the location of the target.

Eye movements were monitored using a desktop-mounted Eyelink 1000 (SR Research, Ottawa, Canada). Stimulus presentation was controlled using Psych-toolbox (Brainard (1997)) and Eyelinktoolbox (Cornelissen, Peters, and Palmer (2002)) for Matlab and run on an Apple powermac. All search and detection arrays were 1024×1024 pixels and displayed on a 25" Sony Trimaster EL OLED monitor with linear gamma. The viewing distance was controlled by use of a chin rest placed 57cm away from the display monitor. At this distance, one pixel is approximately $0.014°$ of visual angle; images subtend $14.3°$ and the targets subtend $0.2°$ of visual angle.

2.1.3. Set-up: Target Detection

Trials were presented in a random order, with a break every 50 trials. Each trial started with a central fixation cross which the observer was required to fixate for 1000ms before the texture appeared. The texture was presented for 200ms, during which time the observer had to maintain a central fixation (within $1.4°$ pixels of the centre of the screen). After the stimulus display period a white noise mask was displayed for 500ms, followed by a blank grey response screen. Observers were informed that there was a target on nearly every trial and their task was simply to press a button to indicate if they had seen the target or not. If observer broke central fixation while the stimulus was displayed, the trial was terminated.

There were eight equally spaced target eccentricities ($r \in [0.70°, 5.57°]$ from the centre of the screen), and eight equally spaced directions. There were nine trials for every possible target location and surface roughness, and an additional

6

258 catch (target absent) trials. This gave a total of 1986 trials[1]. Observers were given feedback only on catch trials: a green screen if they correctly responded that they could not see the target, and a red screen if they responded that they could see the target.

2.1.4. Set-up: Visual Search

Participants were instructed to search for the target and press a key when they found it. The search display was presented until response, or timed out after a minute. There were 70 target present trials for each value of β and the target was positioned randomly with the constraint that it was at least $1.25°$ away from the edge of the surface texture, and not contained in a $2.5° \times 2.5°$ pixel window positioned on the stimulus centre. An additional ten target absent catch trials were included for each surface roughness. This gave a total of 240 trials.

On catch trials, the search display remained on for a fixed time (30, 15 or 5 seconds depending on surface roughness) and observers were given feedback on these trials: a red screen if they responded that they could see the target, and a green screen if they correctly searched the stimulus for the target for the full display period. A task in which the target was (nearly) always present was used in this phase because in previous work using these stimuli (Clarke et al., 2008) observers frequently failed to find targets on rougher surfaces when they had the option of making a target absent response.

2.1.5. Analysis

Statistical analysis was carried out using the `lme4` package for `R`. *p*-values were obtained using the `Anova` function from the `car` package (based on type II Wald χ^2 tests) (Fox and Weisberg, 2010).

2.2. Results

2.2.1. Target Detection

Very few trials were rejected due to failure to maintain central fixation (mean 0.6% rejected trials per participant). These trials are excluded from all further analysis. First we checked each participant's accuracy (Figure 2). We can see that the false positive rate on catch trials is low, ranging from 0.39% to 14.46% with a median of 3.88%. This suggests the accuracy rate on target present trials is a valid measure of target visibility, although even a small number of false alarms suggests we may be potentially over-estimating the probability of finding the target by a modest amount. This is important to note, but it is also critical to keep in mind that it will bias both the stochastic and optimal models to the same extent. (We will return to the issue of false alarms and estimating target visibility in the discussion.)

---

[1]Participant 1 did 1799 trials, and Participant 7 did 2439 trials. These were the first two participants to carry out this part of the experiment, and adjustments to the number of trials was made to make the experiment last approximately 1.5 hours.
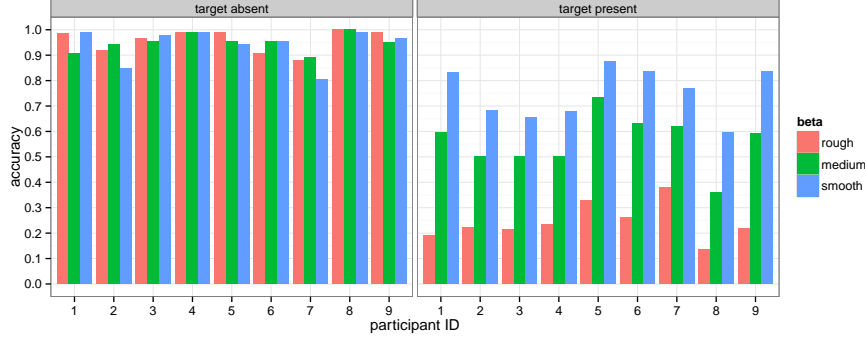
Figure 2: Accuracy of responses to the target for each observer in the target detection study. We can see that the false positives rate is low.

We analysed the results using a general linear mixed model (family=binomial), fit with the model specified as $y \sim \beta * x^2 + y^2$. (It does not make sense to fit $\beta : x^2$ and $\beta : y^2$ interactions - fitting just one of them is enough to allow for the shape of the target detection function to vary with surface roughness.) The model allowed for random slopes for β as well as random intercepts. We find statistically significant effects ($p < 0.001$) of β ($\chi^2(2) = 233.8$), x^2 ($\chi^2(1) = 800.2$) and y^2 ($\chi^2(1) = 1399.2$). The $\beta : x^2$ interaction was also significant ($\chi^2(2) = 13.7$, $p = 0.001$). A simplified version of the target detection function is showing in Figure 3.

2.2.2. Visual Search

Mean accuracy was 99% and 94% for target present and absent trials respectively. We analyse log reaction time (RT) with a linear mixed model: $rt \sim \beta + (\beta|\texttt{participantID})$. As expected, we find that β has a statistically significant effect on log reaction times: $\chi^2 = 243.75$, $p < 0.001$, with slower RTs to find targets on rougher surfaces. In the following section, we will investigate the extent to which the number of fixations required to find the target can be explain by two different models.

3. Search Strategies

Here we describe two different models - stochastic and optimal search - of the search task described in Section 2. Both models are based on the visibility map derived in Section 2.2.1. That is, the probability of detecting a target located at (x, y), when fixating $f_i = (x_f, y_f)$ is given by

$$p(d_{x,y}|f_i, \beta) = F(\vec{b_1} \cdot \vec{\beta} + \vec{b_2}\vec{\beta}(x - x_f)^2 + b_3(y - y_f)^2) \tag{1}$$

where $\vec{\beta}$ is a vector encoding the categorical factor $\beta \in \{\texttt{rough, medium, smooth}\}$ and $\vec{b_1}, \vec{b_1}, b_3$ are the models parameters fit to logistic regression (There is no
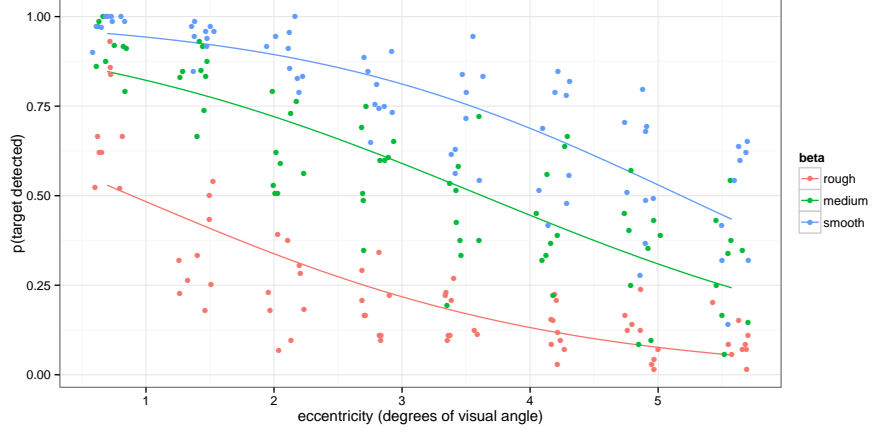
Figure 3: Probability of detecting the target, collapsed over angle (i.e., only taking the target's eccentricity into account). The individual points show each participant's performance, while the lines show a binomial fit (general linear model).

need to include an interaction between β and y.) F is the logistic transform:

$$F(z) = \frac{1}{1 + e^{-z}} \tag{2}$$

This model is fitted to the results of the target detection experiment in Section 2.2.1, collapsing over participants. The coefficients are given in Table 1 and this function is illustrated in Figure 4.

3.1. Stochastic Searcher

The main aim of this study is to explore the extent to which visual search performance can be explained by a random walk. For each fixation, the stochastic searcher uses the target detection function given in Equation 1 to determine
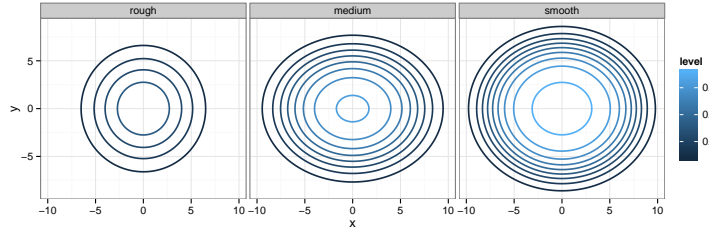

Figure 4: Contour plot showing target detection model.

β	rough	medium	smooth
intercept	-0.17	1.27	2.39
x^2	-0.089	-0.062	-0.074
y^2	-0.103	-0.103	-0.103

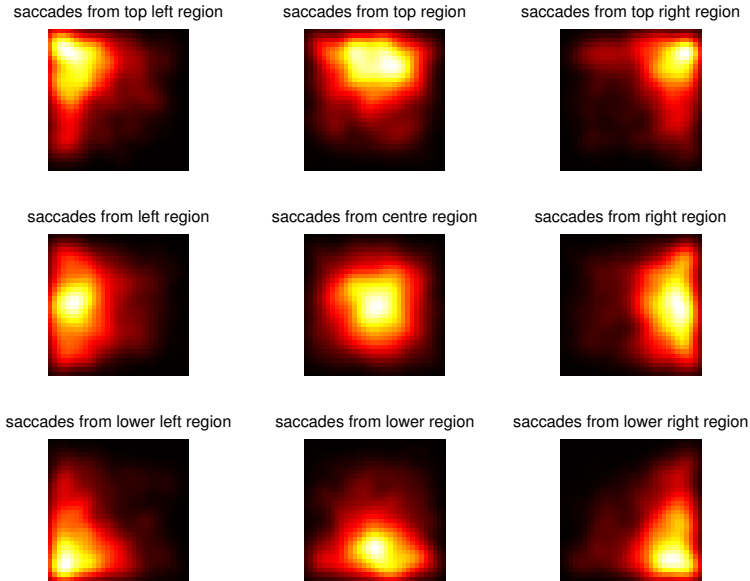Table 1: Coefficients used in Equation 1, the target detection function.

Figure 5: Each subplot shows a hotspot map of fixation locations from a different region of the stimuli. For example, we can see that saccades originating from the corner regions tend to be directed back towards the centre, or along one of the edges.

if the target is present. To decide where to fixate next, this model samples a saccade at random (from empirical data), conditioned on the current fixation location, $S(x', y', x_i, y_i) = p\left((x', y')|(x_i, y_i)\right)$. We base S on the distribution of saccades recorded in Section 2.2.2. To estimate S, we start by quantising the fixations to a $Q \times Q$ grid ($Q = 32$). Then we simply count the number of saccades from (q_{x_1}, p_{y_1}) to (q_{x_2}, p_{y_2}), $1 \leq x_i, y_i \leq Q$. As the last saccade in each trial is likely to be directed towards the target, rather than searching for the target, these saccades are not included in this distribution. We then convolve S with a four-dimensional Gaussian filter ($\sigma = 3$). Figure 5 shows a simplified version of this distribution (with $Q = 3$).

The use of this distribution allows the stochastic searcher to act as a realistic baseline: it will make saccades with amplitudes and directions similar to those made by human observers, and it avoids making saccades to locations outside of the search area. On the other hand, as the probability of making a fixation to a given location is only conditioned on the sample of saccades to that region in previous data, it has no memory of where it has looked before, and it has

no notion of inhibition of return or saccadic momentum. Furthermore, the stochastic searcher does not adjust its behaviour based on the difficulty of the search task or the probability of detecting the target.

3.2. Optimal searcher

The optimal model uses the target detection function, $p(d_{x,y}|f_i, \beta)$, to estimate $p(t_{x,y})$, the probability that the target is located at (x, y), defined as:

$$P_i(x, y) = p(t_{x,y}|\beta, f_1, \ldots, f_i) \tag{3}$$

3.2.1. Probability that target is located at (x, y)

We start off by setting P_0 to our prior on where the target is likely to be located. We will use a uniform distribution over the search area: $P_0(x, y) \propto \hat{P}_0 = 1$. (i.e., $P_0(x, y) = \frac{1}{N^2}$, where N^2 gives the area of the search stimulus). However, P_0 could be set to any arbitrary distribution.

After each fixation, we use Equation 1 to check if the target has been detected. If the target is detected then $P_i = 1$ at the target's location, 0 everywhere else and the search process is terminated. If the target is not detected then, assuming (i) perfect memory and (ii) that $p(d_{x,y}|\beta, f_i)$ $\forall i$ are independent, we define $P_i \propto \hat{P}_i$ as:

$$P_i(x, y) \propto \hat{P}_i(x, y) = p(t_{x,y}|\beta, f_1, \ldots, f_i) \tag{4}$$
$$= p(t_{x,y}|\beta, f_1, \ldots, f_{i-1}) \times p(t_{x,y}|\beta, f_i) \tag{5}$$
$$= \hat{P}_{i-1}(x, y) \times (p(\neg d_{x,y}|\beta, f_i)) \tag{6}$$

The logic of this is as follows. Consider our first fixation $f_1 = (x_1, y_1)$, a potential target location (x', y') close to the current fixation, and let us assume that we have not yet detected the target. If the target was located at (x', y'), then we would have a relatively good chance of detecting it. For the purpose of this explanation, let us assume that $p(d_{x',y'}|f_1) = 0.8$, that is, if we fixate at (x_1, y_1) then we have a 80% chance of detecting the target if it is located at (x', y'). Now let us consider the probability that the target is indeed located at (x', y'). If there were an 80% chance of target detection, but no target was detected, there is (up to a normalisation step to take into account the total number of possible target locations) a 20% chance that the target is located at (x', y'). More generally:

$$p(t_{x',y'}|f_1) \propto p(\neg d_{x',y'}|f_1) \tag{7}$$

We can iterate over successive fixations using Equation 6.

3.2.2. Choosing where to fixate next

Once we have calculated $\hat{P}_i$ we use it to select the next fixation location. As we are only calculating P_i up to a normalisation constant, the map that
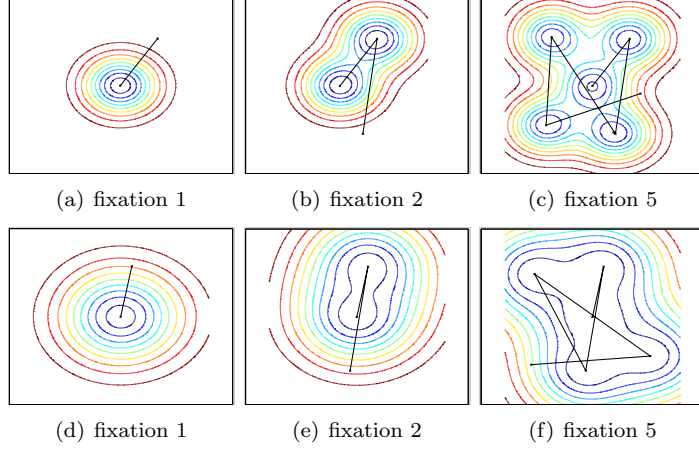
(a) fixation 1   (b) fixation 2   (c) fixation 5

(d) fixation 1   (e) fixation 2   (f) fixation 5

Figure 6: This figure shows how P_i is updated with each fixation and used to plan the next fixation. (a), (b) and (c) show fixation 1,2, and 5 in the hard condition, while (d), (e) and (f) show how search progresses in the easy condition.

we actually calculate, $\hat{P}_i = \hat{P}_{i-1}(x,y) \times p(\neg d_{x,y}|\beta, f_i)$, has the property that $\hat{P}_i(x,y) \le \hat{P}_{i-1}(x,y) \; \forall i, x, y$. This can be thought of as measuring how much we still have to search different regions of the stimulus. Our model will make a fixation at the point $f_{i+1} = (x_{i+1}, y_{i+1})$ that maximises the difference between P_i and P_{i+1}. This can be expressed as:

$$(x_{i+1}, y_{i+1}) = \underset{f_j=(x_j,y_j)}{\arg\min} \sum_{x,y} P_i(x,y)p(\neg d_{x,y}|\beta, f_j) \tag{8}$$

This process is illustrated in Figure 6. In order to reduce computation time, the simulation takes place at a lower resolution, with $N = 256$ rather than 1024.

3.3. Comparing human observer to the models

We now compare the two models to human performance, as detailed in Section 2.2.2 by simulating search over the same number as trials used in the human experiment. In agreement with Najemnik and Geisler (2008), we find that, in terms of the number of saccades required to find the target, human performance is consistent with an optimal search strategy. However, as we can see from Figure 7, the stochastic searcher offers very similar performance. Indeed, a one-tailed, paired t-test between the two models gives $t(179) = 0.70$, $p = 0.24$.

We can also look at the saccade statistics to see if they offer a way to distinguish between the two models and determine which offers the more human-like behaviour. Figure 8 shows some example scanpaths. We can see that the optimal model favours making fixations in the centres of the four quadrants of the search area. These points maximise the area of the stimulus in the effective
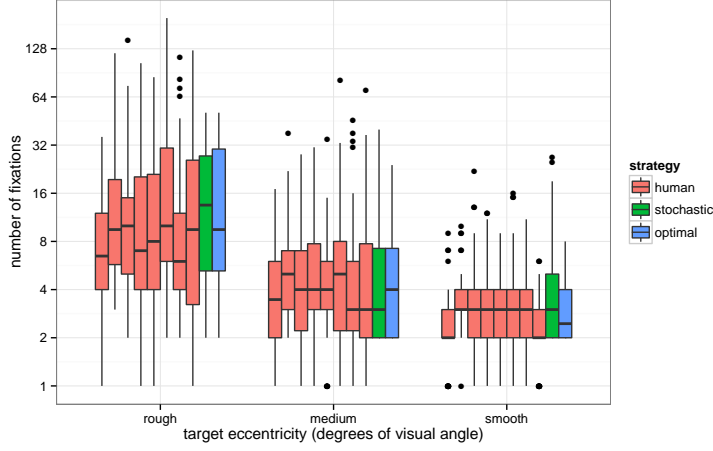
Figure 7: Number of saccades made by human observers and search simulations. The two models differ very little from human observers.

field of view while minimising the overlap with earlier fixations made to other regions. While this involves re-fixating a previously fixated location, given the low probability of detection on rough surfaces (Figure 3), this is not as unintuitive as it may initially appear - there is a 50% chance that the model will miss a target at fixation on a rough surface. We can also see that after fixating the centre of each quadrant three times (giving a probability $p \approx 1 - 0.5^3 = 0.875$ of detecting the target at these points), the model switches to fixate the in-between regions.

Figure 9 shows the distribution of saccade amplitudes and directions. There are large differences between the behaviour of the optimal model and human observers.

Interestingly we see that even though the stochastic model is only constrained by the distribution of saccades made by human observers, this is sufficient to give a good match for the displacement over two saccades, and hence it exhibits a similar level of inhibition of return as the human observers.

3.4. Discussion

In general, the stochastic model results provide a good match for human behaviour, both in terms of the number of fixations to find the target and the overall pattern of fixations. Our optimal model finds the target in a similar number of fixations, but does not match well with human saccade behaviour on any other metrics. This conclusion is at odds with Najemnik and Geisler (2008), who pointed to the similarity in fixation distributions between human and ideal observers as evidence that humans employ an optimal strategy. Specifically, the optimal observer predicts a doughnut shaped distribution with peaks above and below the central fixation cross, which they also observed in their human
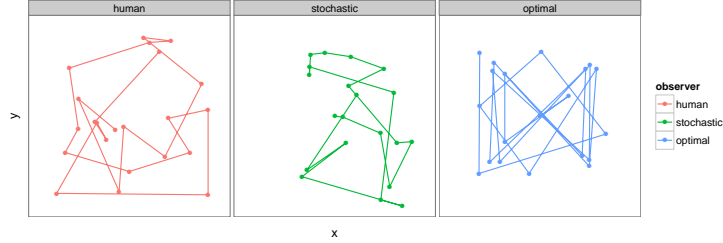
13

Figure 8: Example scanpaths made by a human, stochastic and optimal observer for a rough surface. We can see that similarly to Najemnik and Geisler (2008), the optimal model shows a preference for several fixed points on the stimulus. These are the locations which maximise the information gained via para-foveal vision, and minimise the proportion of the effective field of vision that falls outside of the search area. The centre is avoided after the initial fixation because it is frequently sampled by peripheral vision.

searchers. One plausible reason our results from nine observers did not produce this pattern is that our experiments used a square search array while their array was circular. In order to investigate this potential reason for the discrepant findings, we repeated our visual search experiment with a circular search array with three of our observers. The resulting hotspot maps for each observer for square and circular search stimuli are shown in Figure 10. The distinctive pattern of fixation locations observed in Najemnik and Geisler (2008) was not replicated.

4. Reanalysis of Clarke et al. (2009)

In the previous experiment we found evidence that a random walk is a viable explanation of human visual search performance. In this section we apply the same analysis to a similar existing dataset (Clarke et al., 2009).The empirical data were collected in a different lab with a different eye-tracker and fixation filter, and should therefore provide a good test of the robustness of our results. In addition, in the visual search task used in this dataset the target's eccentricity is systematically varied, allowing us to compare the stochastic and optimal models to human performance separately for different target eccentricities. As in Experiment 1, we first generated a target visibility function using a target detection task. We collected target visibility data from two naive observers who did not perform the visual search task.

4.1. Target detection experiment

4.1.1. Observers

Two observers carried out all the trials, split into twenty blocks of 132 trials each, over a number of days. They were paid £50 each. The research was
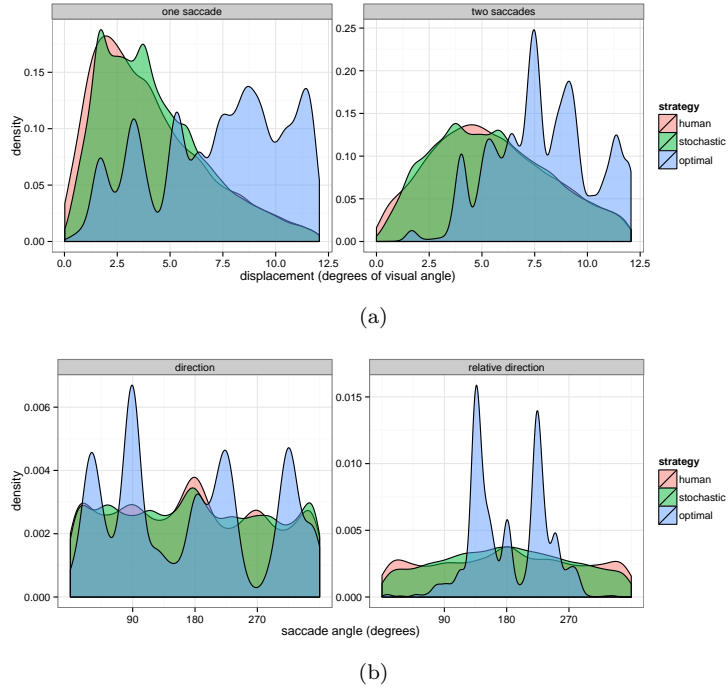
(a)

(b)

Figure 9: (a) Displacement over one and two saccades. (b) Absolute and relative direction of saccades. The fact that the stochastic model matches the distribution of displacement and direction over a single saccade is a direct result of sampling from empirical data and is therefore unsurprising. The distributions over two saccades are, however, more interesting: we see that there is no need to add an IOR or saccade momentum component to the stochastic model beyond what is implicitly included in the distribution of saccades from which the model samples.

(a) square search areas.
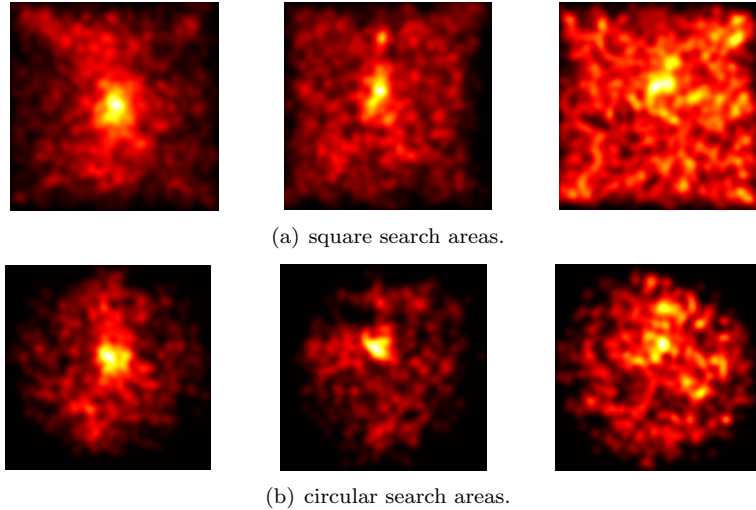

(b) circular search areas.

Figure 10: Hotspot maps for square search areas for participants 2, 4 and 5 (left to right). We can see some individual differences, in particular, how strong the central bias is for each observer. However, there is no clear tendency to fixate above and below the fixation.

conducted in accord with the Code of Ethics of the World Medical Association (Declaration of Helsinki) and informed consent was obtained from both observers.

4.1.2. Stimuli

Surface textures were created as detailed above. For the target present trials, the target was located at one of 72 potential locations: nine different eccentricities were used ($0.84° \leq r \leq 7.5°$), and eight evenly-spaced orientations. For each parameter combination, 20 different trials were created. Based on pilot results, we created 160 target absent trials for each value of β, giving a total of 2160 target present trials and 480 target absent trials. This ratio of target present to absent trials ensured that observers made roughly equal numbers of "present" and "absent" responses.

4.1.3. Set-up

Observers were instructed to keep their eyes fixated on the centre of the image. After each trial, they were asked to respond with a button press to indicate if they had seen the target or not. Each trial consisted of a fixation cross (500ms), stimulus (200ms), white noise mask (500ms), and finally a fixation cross was displayed until a target present or absent response was given.

A Tobii x50 eyetracker was used to sample the observers' gaze every 20ms and trials were included in the further analysis only if (i) the mean gaze location was within 1° of the central fixation cross (ii) the standard deviation of the gaze's x and y components was less than 2/3°.

16

4.1.4. Results

13.6% of trials were removed from analysis (due to breaking central fixation). We first checked the false positive rate on target absent trials, which were similar to those seen in Experiment One (less than 10% in all conditions for both observers). We therefore focus on the probability of detecting the target when it was present as a measure of visibility. We collapse over the two observers and fit a model as in Section 2.2.1. Results and model coefficients are similar.

4.2. Visual Search Dataset

The visual search data were taken from an experiment originally published in Clarke et al. (2009). The methods and data are summarised here.

4.2.1. Observers

Seven observers, aged 18-30, were given several practice trials and were informed that the target would be present in all trials and would always be an indent in the surface of the same size and shape. They were instructed to respond by pressing the space bar on the keyboard once they had found the target. No time limit was imposed on the task. Observers were told to inform the experimenter if they were having great difficulty in finding the target, in which case they were allowed to skip the trial (in practice this accounted for less than 1% of trials).

4.2.2. Surface Stimuli

This stimuli were created as in Section 2.1.2. For each trial a target was positioned randomly on a circle, centred on the middle of the image, with radius $1.7° \pm 0.7°$, $3.8° \pm 0.7°$ or $5.9° \pm 0.7°$ visual angle.

4.2.3. Set-up

Stimulus presentation was controlled by Clearview (Tobii Technology Inc). All stimuli were 1024×1024 pixels in size and displayed on a NEC LCD2090UXi monitor. The pixel dimensions were 0.255mm by 0.255mm resulting in images with physical dimensions 26.1cm by 26.1cm. The monitor was linearly calibrated with a Gretag-MacBeth Eye-One; maximum luminance was set at 120 cd/m^2. This results in the rendered images appearing as if they were under bright room lighting conditions.

A Tobii x50 eye-tracker was used to record observers' gaze patterns. The fixation filter was set to count only those fixations lasting longer than 100 ms within an area of 30 pixels. The accuracy of the eye-tracker was $0.5°$ to $0.7°$ and the spatial resolution was $0.35°$. The viewing distance was controlled by use of a chin rest placed 87cm away from the display monitor. At this distance, one pixel is approximately 1 arcminute of visual angle; images subtend $16.7°$ and the targets subtend $0.66°$ of visual angle.

4.2.4. Results

The number of fixations required to find the target is shown in Figure 11. Using a generalised linear mixed model (`family="poisson"`) we confirm that both roughness (β) and target eccentricity, along with their interaction, have a statistically significant effect ($p < 0.05$) on the number of fixations required to detect the target.

4.3. Comparison with Search Models

We now compare the model to human performance, as detailed in Section 4.2.4 by simulating search over the same number as trials as used in the original experiment. The results closely match those of Experiment 1. In terms of the number of saccades required to find the target, we find the stochastic searcher offers very similar performance to optimal model and to the seven human observers over the range of surface roughness's and target eccentricities used (Figure 11). A one-tailed, paired t-test between the two models gives $t(359) = -0.25$, $p = 0.6^2$.

5. General Discussion

Human search behaviour can be closely modelled by a stochastic process, not only in the number of fixations required to find the target, but also in more detailed measures of search behaviour, such as the spatial distribution of fixations, the frequency of saccade amplitudes, and the relative angles of sequences of saccades. The stochastic model also makes a similar number of re-fixations.

As with Najemnik and Geisler (2005, 2008), we find that an optimal model requires a similar number of fixations to detect the target as human observers. However, our optimal model produces scan paths and saccade statistics that differ substantially from both the human and stochastic model behaviour. It is interesting that the stochastic and optimal models – two models that have very different architectures and produce very different search behaviour – could both take a similar number of fixations to detect the target. Clearly there is more than one way to achieve this single end, and maximising the probability of detecting the target separately for each fixation is not a requirement for relatively efficient search to be achieved. It should be mentioned that Najemnik and Geisler (2005) did compare their optimal model to a random baseline, and that this baseline made far more fixations to detect the target than either the optimal model or human searchers. However, their random model selected random coordinates within the search stimulus as the target of each fixation in the sequence. This baseline differs from the current model in that it does

---

[2]As in Experiment 1, this difference becomes statistically significant if more data is generated from the simulations: with 10 times as many trials we get $t(3599) = 2.6$, $p = 0.005$ As before, this difference corresponds to the optimal model finding the target on average half a fixation faster than the stochastic model.
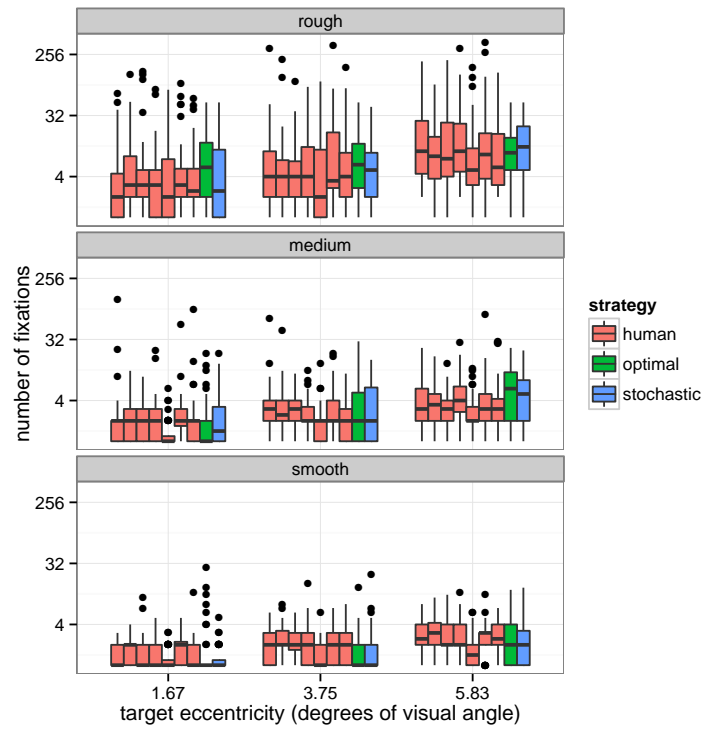
Figure 11: Number of saccades made by human observers and search simulations. The two models differ very little from human observers. We can see that the stochastic search model is sufficient to explain the number of fixations required to find the target.

not take into account the natural tendencies in saccade behaviour that make some fixations and sequences of fixations more likely than others. Our results demonstrate that these tendencies alone, irrespective of any knowledge about previous fixations or target detection probabilities, can produce efficient search.

Although our optimal model follows a similar logic as Najemnik and Geisler (2005, 2008), it differed in a number of details. Our focus in this study is on developing the stochastic model and comparing it to human behaviour, and for this, the question of whether or not is it possible for an optimal model to produce saccade behaviour that is a better match for human behaviour is not a central concern. That said, it is worth noting some of the key differences between the studies. One important difference between our optimal search model and the one presented in Najemnik and Geisler (2008) is that our model is essentially deterministic and does not include fixation-to-fixation internal noise. Another is that we used textured surfaces instead of noise as a search display. Finally, we used a larger group of naive observes while the observers in Najemnik and Geisler (2008) were the two authors of the paper. A great deal of practice with the search display together with detailed knowledge about optimality could influence search strategies and produce behaviour that is better matched to an optimal pattern.

Despite these differences, our modeling results are consistent with Najemnik and Geisler (2005, 2008) in matching human performance in terms of the number of fixations to find the target. However, a key difference is that Najemnik and Geisler (2008) report a tendency to fixate above and below the centre of the search display for both humans observers and their optimal model. This tendency did not appear in our human data, even when we re-ran the search experiment using a circular search display to match the one used in Najemnik and Geisler (2008).

A final important difference between our study and Najemnik and Geisler (2005, 2008) is in the experiment used to collect data to create the target visibility map. Najemnik and Geisler (2005, 2008) made use of a 2AFC procedure in which observers had to select which of two intervals contained the target. Target location was blocked and observers were also spatially cued to the target location on each trial. Cueing a target's location has long been known to improve performance (Posner, 1980), specifically by enhancing the target signal (Yeshurun and Carrasco, 1998). We therefore opted to use a simpler target detection procedure in which the observer did not know the target's location ahead of time. We chose this because it is a better approximation of visual sensitivity in the context of visual search, where the target location is also not known. The downside of this approach is that the spatial distribution of false positives is unknown and hence calculating d' is problematic. In this study we minimised false positives by lowering the number of target absent trials, and then used accuracy on the target present trials as a measure of sensitivity. While this potentially over-estimates human performance, the same criticism can be made of Najemnik and Geisler's 2AFC task. Overall, it seems unlikely that modest changes to the target visibility map will have much effect on the optimal search model's behaviour, and in any case, the same map is used in both the stochastic

and optimal model so small changes to the sensitivity function will increase and lower the number of fixations required to detect the target to a similar degree for both models.

We do not wish to claim that stochastic selection of saccades is the only process involved in search. As stated in the introduction, search strategy and feature extraction work together to produce search behaviour. A stochastic process would work in concert with guided search in a more typical search context in which there are many objects and/or contextual information available. One could imagine that if there are several search items that could potentially be the target, a random walk model could be used to choose which item should be fixated next. Also, although our model did not need any form of memory or inhibition of return to achieve human-like behaviour, we do not mean to suggest there is no inhibition of return in human search. Indeed, as our stimuli contain no search objects, any IOR processes would have to be operating in spatiotopic coordinates defined with respect to stimulus boundaries, rather than being applied to discrete search objects. IOR is strengthened by objects (Jordan and Tipper, 1998), and this may be particularly true when the eyes move, necessitating inhibition of spatiotopic, rather than retinal, coordinates (e.g., Krüger and Hunt (2013)).

Our results suggest that the process of deciding where to look next may be driven by a simple random selection from a subset of possible saccades. What determines this subset of possible saccade is an interesting question, but is likely to involve a combination of visual and motor constraints together with a lifetime of experience searching for objects. The mechanism that has been assumed to subserve efficient eye movement behaviour in natural tasks is reinforcement learning (e.g. Hayhoe and Ballard (2014)). Prior learning during search tasks could cause particularly effective saccades and sequences of saccades to be selected rapidly and efficiently, without the need for a computationally taxing process of keeping track of the probability of a target being in any possible location given target visibility and a memory of all previous fixations over an extended sequence. Our conclusion is consistent with recent work suggesting humans are sub-optimal in search (Morvan and Maloney (2012); Verghese (2012); Zhang et al. (2012)), in that saccades during search do not appear to maximise the probability of detecting the target based on previous fixations and knowledge about the limits of our own visual acuity. Fortunately, our results suggest this kind of optimality is not a pre-requisite for search efficiency: A stochastic model based on human saccade data can perform as well as an optimal model.

Acknowledgements

Brainard, D. H., 1997. The psychophysics toolbox. Spatial vision 10, 433–436.

Clarke, A. D. F., Green, P. R., Chantler, M., 2009. Modelling visual search on a rough surface. Journal of Vision 9x, 1–12.

Clarke, A. D. F., Green, P. R., Chantler, M., Emrith, K., 2008. Visual search for a target against a $1/f^\beta$ continuous textured background. Vision Research 48, 2193–2203.

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 (4), 613–617.

Fox, J., Weisberg, H. S., 2010. An R companion to applied regression. Sage Publications.

Gao, D., Mahadevan, V., Vasconcelos, N., 2008. On the plausibility for the driscriminant center-surround hypothesis for visual saliency. Journal of Vision 8(7) (13), 1–18.

Gilchrist, I. D., Harvey, M., 2006. Evidence for a systematic component within scan paths in visual search. Visual Cognition 14, 704–715.

Hayhoe, M., Ballard, D., 2014. Modeling task control of eye movements. Current Biology 24 (13), R622–R628.

Hwang, A. D., Higgins, E. C., Pomplun, M., 2009. A model of top-down attentional control during visual search in complex scenes. Journal of Vision 9(5), 1–18.

Itti, L., Baldi, P. F., 2009. Bayesian surprise attracts human attention. Vision Research 49, 1295–1306.

Itti, L., Koch, C., May 2000. A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Research 40 (10-12), 1489–1506.

Jordan, H., Tipper, S. P., 1998. Object-based inhibition of return in static displays. Psychonomic Bulletin & Review 5 (3), 504–509.

Krüger, H. M., Hunt, A. R., 2013. Inhibition of return across eye and object movements: The role of prediction. Journal of Experimental Psychology: Human Perception and Performance 39 (3), 735.

Land, M. F., McLeod, P., 2000. From eye movements to actions: how batsmen hit the ball. Nature neuroscience 3 (12), 1340–1345.

Morvan, C., Maloney, L. T., 2012. Human visual search does not maximize the post-saccadic probability of identifying targets. PLoS computational biology 8 (2), e1002342.

Motter, B. C., Holsapple, J. W., 2001. Separating attention from chance in active visual search. In: Braun, J., Koch, C., Davis, J. (Eds.), Visual attention and neural circuits. Cambridge, MA: MIT Press, pp. 159–175.

Najemnik, J., Geisler, W. S., 2005. Optimal eye move ment strategies in visual search. Nature 434, 387–391.

Najemnik, J., Geisler, W. S., 2008. Eye movement statistics in humans are consistent with an optimal search strategy. Journal of Vision 8(3), 1–14.

Over, E. A., Hooge, I. T., Erkelens, C. J., 2003. Visual search: saccade parameters depend on the shape of the search area [abstract]. Journal of Vision 3 (429), 251–261.

Over, E. A., Hooge, I. T., Vlaskamp, B. N., Erkelens, C. J., 2007. Coarse-to-fine eye movements strategy in visual search. Vision Research 47, 2272–2280.

Pomplun, M., 2007. Advancing area activation towards a general model of eye movements in visual search. In: Gray, W. D. (Ed.), Integrated Models of Cognitive Systems. New York Oxford University Press, pp. 120 – 131.

Pomplun, M., Shen, J., Reingold, E. M., 2003. Area activation: A computational model of saccadic selectivity in visual search. Cognitive Science 27, 299 – 312.

Posner, M. I., 1980. Orienting of attention. Quarterly journal of experimental psychology 32 (1), 3–25.

Rao, P. N., Zelinsky, G. J., Hayhoe, M. M., Ballard, D. H., 2002. Eye movements in iconic visual search. Vision Research 42, 1447–1463.

Reeves, A., Santhi, N., DeCaro, S., 2005. A random-ray model for speed and accuracy in perceptual experiments. Spatial Vision 115, 73–83.

Rutishauser, U., Koch, C., 2007. Probabilistic modeling of eye movement data during conjunction search via feature-based attention. Journal of Vision 7 (5), 1–20.

Stone, M., 1960. Models for choice-reation time. Psychometrika, 251–260.

Tavassoli, A., van der Linde, I., Bovik, A. C., 2009. Eye movements selective for spatial frequency and orientation during active vision search. Vision Research 49, 173–181.

Verghese, P., 2012. Active search for multiple targets is inefficient. Vision research 74, 61–71.

Wolfe, J. M., 2007. Guided search 4.0: Current progress with a model of visual search. In: Gray, W. (Ed.), Integrated Models of Cognitive Systems. New York: Oxford, pp. 99–119.

Yeshurun, Y., Carrasco, M., 1998. Attention improves or impairs visual performance by enhancing spatial resolution. Nature 396 (6706), 72–75.

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

Zhang, H., Morvan, C., Etezad-Heydari, L.-A., Maloney, L. T., 2012. Very slow search and reach: Failure to maximize expected gain in an eye-hand coordination task. PLoS computational biology 8 (10), e1002718.