Human search for a target on a textured background is consistent with a stochastic model

Alasdair D. F. Clarke

School of Psychology, University of Aberdeen, Aberdeen, UK

Patrick Green

School of Life Sciences, Heriot-Watt University, Edinburgh, UK

Mike J. Chantler

School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, UK

Amelia R. Hunt

School of Psychology, University of Aberdeen, Aberdeen, UK

Previous work has demonstrated that search for a target in noise is consistent with the predictions of the optimal search strategy, both in the spatial distribution of fixation locations and in the number of fixations observers require to find the target. In this study we describe a challenging visual-search task and compare the number of fixations required by human observers to find the target to predictions made by a stochastic search model. This model relies on a target-visibility map based on human performance in a separate detection task. If the model does not detect the target, then it selects the next saccade by randomly sampling from the distribution of saccades that human observers made. We find that a memoryless stochastic model matches human performance in this task. Furthermore, we find that the similarity in the distribution of fixation locations between human observers and the ideal observer does not replicate: Rather than making the signature doughnut-shaped distribution predicted by the ideal search strategy, the fixations made by observers are best described by a central bias. We conclude that, when searching for a target in noise, humans use an essentially random strategy, which achieves near optimal behavior due to biases in the distributions of saccades we have a tendency to make. The findings reconcile the existence of highly efficient human search performance with recent studies demonstrating clear failures of optimality in single and multiple saccade tasks.

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 prioritizes information both within each fixation and across a sequence of fixations.

Much of the research on visual search to date has formalized this general issue by focusing on questions of *feature extraction* and of *strategy*. Feature extraction includes both top-down guided search (Wolfe, 2007; Zelinsky, 2008) and stimulus-driven (saliency) effects (Gao, Mahadevan, & Vasconcelos, 2008; Itti & Baldi, 2009; Itti & Koch, 2000). For the abstract and discrete search items commonly used as visual-search stimuli, categorical features such as color, 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, & Reingold, 2003; Rutishauser & 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 (Hwang, Higgins, & Pomplun, 2009; Pomplun, 2007; Rao et al., 2002; Tavassoli, van der Linde, & Bovik, 2009; Zelinsky, 2008). 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 eye movements.

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 posteriori (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 represent either a search item or the center 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 (Clarke, Green, & Chantler, 2009; Itti & Koch, 2000; Pomplun et al., 2003; Rao et al., 2002; Rutishauser & Koch, 2007; 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 behavior 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 & 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 center 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 & 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 maximize task-relevant information gain (for a recent review, see Hayhoe & Ballard, 2014).

In contrast with the notion that humans are close to optimal in search behavior is recent evidence of suboptimality 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 judgment about a dot that could appear in either the left- or the rightmost square. When the squares are closely spaced, the center 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 center location is no longer optimal; instead, observers can maximize accuracy by selecting either the left or the 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 they 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 they were given a monetary reward for each correct response. This finding has been recently replicated and generalized by Clarke and Hunt (2015), and a similar conclusion was also 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 et al. (2012), who found suboptimal eye–hand coordination in a reaching task.

How can these demonstrations of suboptimal eye-movement behavior be reconciled with the findings of Najemnik and Geisler (2005, 2008)? It is unlikely that observers would be suboptimal 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 behavior 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 IOR and saccadic momentum. For example, Over et al. (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 that these systematic

tendencies can be hard to detect in scan paths because of interactions with salience-based object selection.

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 increases, it continues to account for a sizable fraction of performance. In natural-scene viewing, spatial biases to move the eyes in particular ways play an important role in fixation selection; and indeed, how the eyes move (in terms of saccade amplitude and direction) can provide a better account of fixation selection than can what visual information is selected (Tatler & Vincent, 2009). Random walks have been successfully used to model an observer's speed and accuracy in present/ absent forced-choice experiments (Reeves, Santhi, & DeCaro, 2005; Stone, 1960). 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 behavior is therefore that natural search behavior 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 (for an example, see Clarke et al., 2008, figure 1). We used textured surfaces because they appear naturalistic but, unlike photographs of natural scenes, are fully controlled and parameterized. 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 that we have already reviewed, 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. The results demonstrate that the stochastic model closely matches the number of fixations required to detect the target in human data.

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 unaware of 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. This map was used to determine, for each fixation generated by the stochastic model, whether or not the target had been found. We therefore designed the target-detection task to match as closely as possible the detection task that observers would need to perform during each fixation while searching for the target. The target was presented at any one of 64 locations, and participants needed to state whether they detected it or not. A small number of catch trials were included to permit an estimate of each participant's false-alarm rate. It is important to note that this is a departure from the method used by Najemnik and Geisler (2005, 2008), in which the target-visibility maps were generated from a two-alternative forced-choice (2AFC) task in which the target location was cued on every trial and then presented on half of the trials. Their method has the advantage of allowing the experimenter to take false positives into account and calculate d'. However, a flaw in this method is that certainty about the target's location allows covert attention to be allocated to the region of the target, potentially increasing visual sensitivity (Yeshurun & Carrasco, 1998). This could lead to an overestimation of visual sensitivity during search, particularly for targets that are more difficult to detect.

Although a detection task with an uncertain target location provides a conceptually better match to the task of finding a target during search, the downside of our method is that false positives may be problematic. While in Najemnik and Geisler's (2005, 2008) work 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 minimize 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.

To determine which method was more sensitive to search difficulty manipulations and less likely to overinflate sensitivity, we also explicitly compared the detection rates across difficulty and eccentricity for these two sensitivity-measurement methods (signal detection at known locations vs. simple detection with location uncertainty). The results of this short pilot experiment are presented in the Appendix. The results from this pilot confirm our hypothesis that cuing the target's location to the observer increases detection rates particularly for difficult search conditions and makes detection performance less sensitive to differences in target visibility, presumably because attention can be deployed to the expected target location. A lack of sensitivity to the difficulty manipulation is particu-
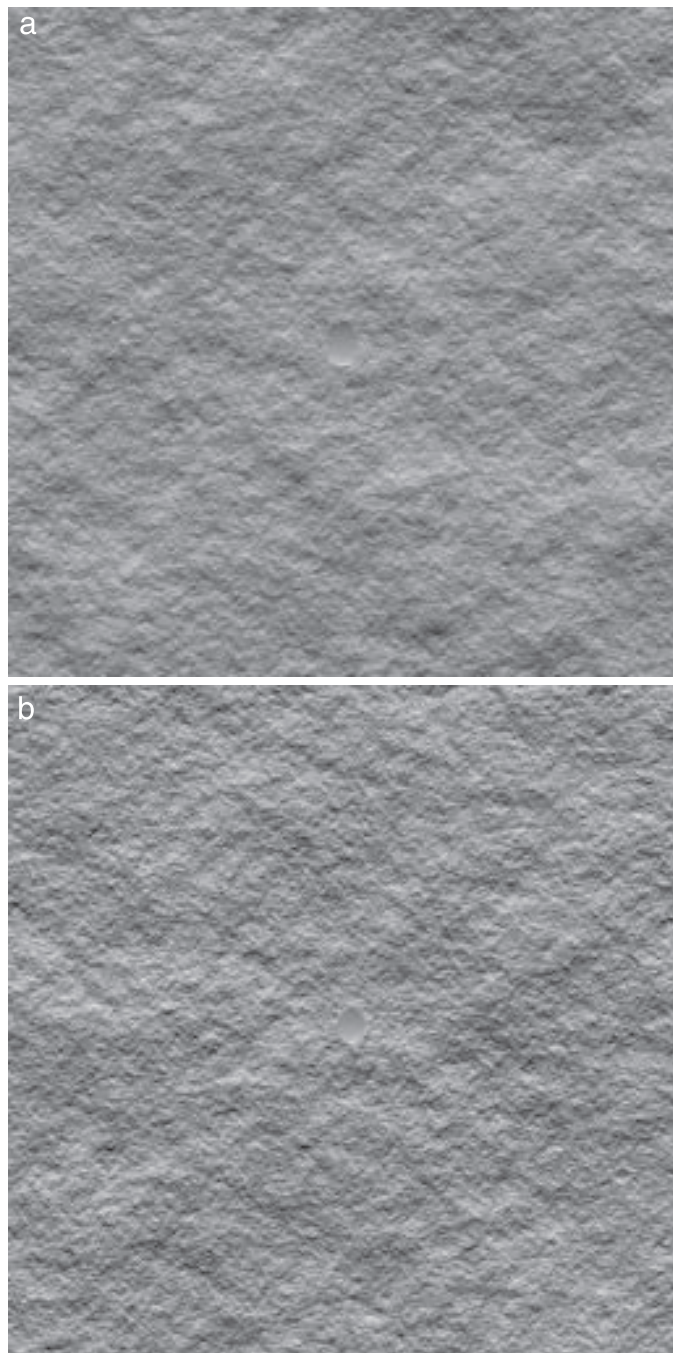
Figure 1. Example stimuli. This is a 256 × 256 pixel crop of one each of the (a) smooth and (b) rough surfaces. In both cases, the target is shown in the center of the image. The stimuli used in the experiment were 1024 × 1024 pixels in size, making the target much smaller relative to the search area than is shown here. The slight differences in the target's shape are due to randomness in the surface at the location of the target.

larly problematic for direct application of the 2AFC measure to the search context, because during search we see clear differences in the search task across difficulty. We therefore conclude that the target-detection task described later is a more appropriate measure than 2AFC for our experiment. 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.

Methods

Observers

Nine observers, 20–29 years old (mean age = 23.8 years), with corrected or corrected-to-normal vision took part in the experiment. All were unaware of the purpose of the study. Two of the participants were undergraduate research assistants 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 min) and £15 for the target-detection part (1 hr 30 min to 2 hr 30 min). 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.

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 at $\sigma_{RMS} = 1.1$. 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.

Eye movements were monitored using a desktop-mounted Eyelink 1000 (SR Research, Ottawa, Canada). Stimulus presentation was controlled using Psychtoolbox (Brainard, 1997) and EyelinkToolbox (Cornelissen, Peters, & Palmer, 2002) for MATLAB and run on an Apple Power Mac. All search and detection arrays were 1024 × 1024 pixels and displayed on a 25-in. Sony Trimaster EL OLED monitor with linear gamma. The viewing distance was controlled by use of a chin rest placed 57 cm away from the display monitor. At this distance, 1 pixel is approximately 0.014° of visual angle; images subtend 14.3° and the targets 0.2° of visual angle.
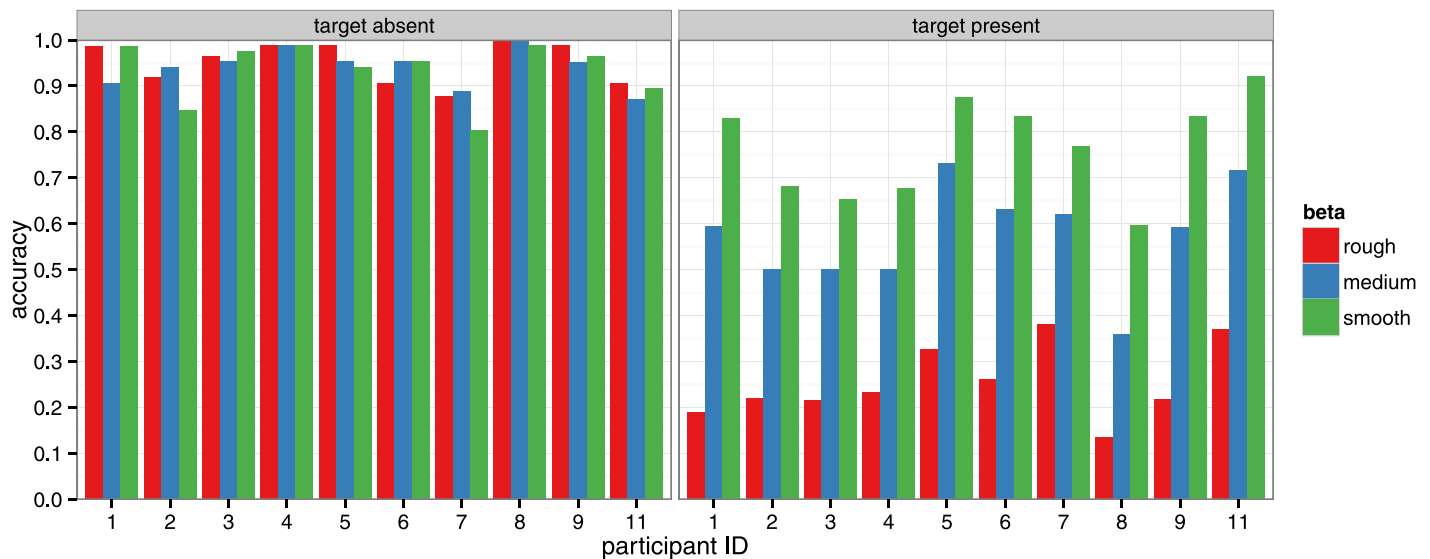
Figure 2. Accuracy of responses to the target for each observer in the target-detection study. The false-positive rate is low.

Setup: 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 1000 ms before the texture appeared. The texture was presented for 200 ms, during which time the observer had to maintain a central fixation (within 1.4° of the center of the screen). After the stimulus display period a white-noise mask was displayed for 500 ms, followed by a blank gray 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 the 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 center of the screen) and eight equally spaced directions. There were nine trials for every possible target location and surface roughness, and an additional 258 catch (target-absent) trials. This gave a total of 1,986 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.

Setup: 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 1 min. 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° window positioned on the stimulus center. An addi-

tional 10 target-absent catch trials were included for each surface roughness. This gave a total of 240 trials. Trials were presented in a random order.

On catch trials, the search display remained on for a fixed time (30, 15, or 5 s, 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.

Analysis

Statistical analysis was carried out using the lme4 (Bates et al., 2013) package for R. The p values were obtained using the Anova function from the car package (based on type II Wald chi-square tests; Fox & Weisberg, 2010).

Results

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 false-positive rate (Figure 2) and found them acceptably low, ranging from 0.39% to 14.46%, with a median of 3.88%. This suggests that accuracy (i.e., hit 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 overestimating the probability of finding the target by a modest amount.

We analyzed the results using a general linear mixed model (family = binomial), fitted with the model specified as $y \sim \beta * (x^2 + y^2)$, where x and y are the
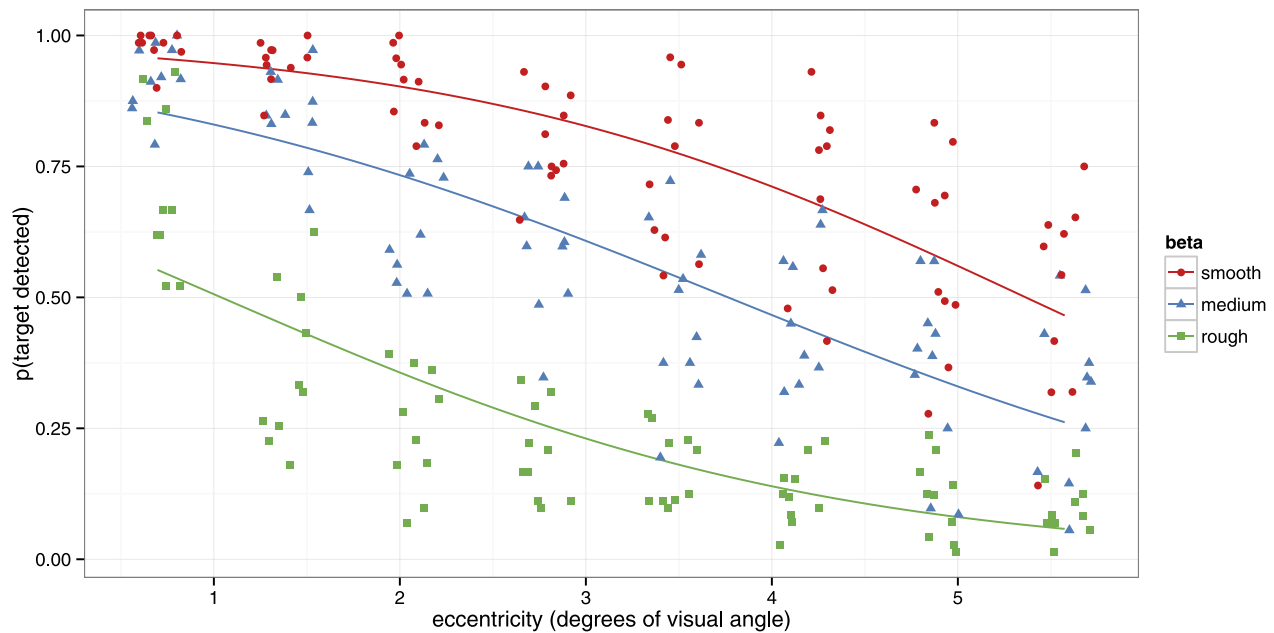
Figure 3. Probability of detecting the target, collapsed over angle (i.e., taking only the target's eccentricity into account). The individual points show each participant's performance, while the lines show a binomial fit (general linear model).

coordinates of the target. The model allowed for random slopes for β as well as random intercepts. We find statistically significant effects ($p < 0.001$) of β, $\chi^2(2) = 3676$; x^2, $\chi^2(1) = 965$; and y^2, $\chi^2(1) = 1735$. The interaction of β and x^2 was also significant, $\chi^2(2) = 20.2$. A simplified version of the target-detection function is shown in Figure 3. The effect of ϕ is illustrated in Figure 4. From this figure, we can see that an ellipsoidal model that is close to radially symmetric will be sufficient to model human performance.

Visual search

Mean accuracy was 99% and 94% for target-present and target-absent trials, respectively. We analyze log reaction time with a linear mixed model: $\log(rt) \sim \beta +$ (β|participantID). As expected, β has a statistically significant effect on log reaction times, $\chi^2 = 243.75$, $p < 0.001$, with longer reaction times to find targets on rougher surfaces. We also examined the spatial distribution of fixation locations using hot-spot maps. As can be seen in Figure 5a, the observers exhibit a central bias (Tatler, 2007) in their fixation locations, although there appear to be individual differences in the variance and spread of fixations. Importantly, the distinctive "doughnut" pattern of fixation locations observed by Najemnik and Geisler (2008)—and used as evidence that human observers are consistent with optimal search strategies—was not replicated. To explore whether this could be explained by the difference in the shape of the stimuli (we used square search areas, while Najemnik and Geisler used circular

areas), we retested a subset of the participants in the same experiment but using circular stimuli. As Figure 5b shows, this had no effect on the pattern of fixations.

Search strategies

Here we describe a stochastic model of the search task described later, based on the visibility map. That is, the probability of detecting a target d located at (x, y), when fixating $f_i = (x_f, y_f)$, is given by

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

where $\vec{\beta}$ is a vector encoding the categorical factor $\beta \in$ {rough, medium, smooth} and $\vec{b_1}$, $\vec{b_1}$, and $\vec{b_3}$ are the model's parameters to be fitted. The function F is the logistic transform:

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

This model is fitted to the results of the target-detection experiment later, collapsing over participants. The coefficients are given in Table 1 and the function is illustrated in Figure 6.

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 if the target is present. To decide where to
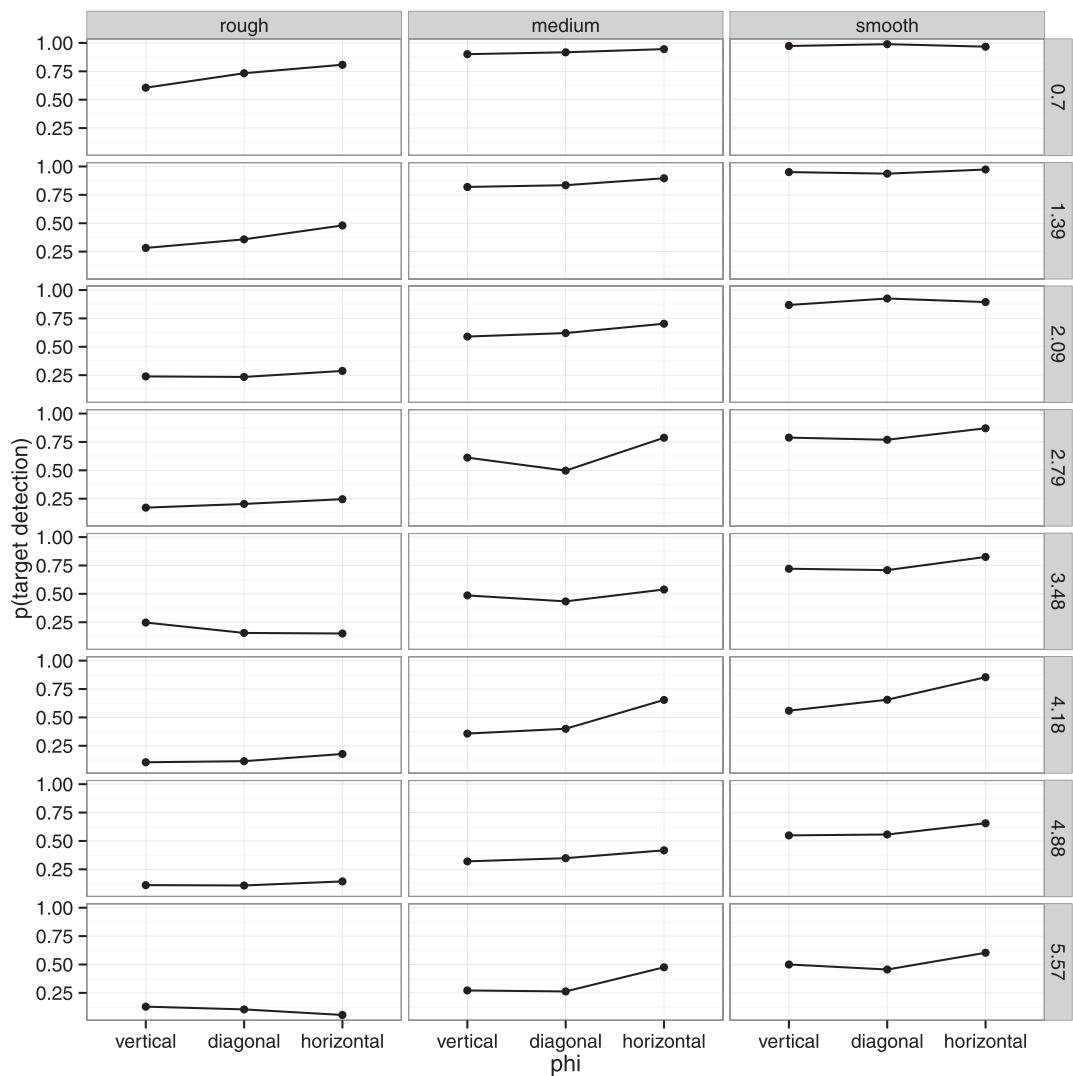
Figure 4. Comparing the probability of detecting the target for horizontal, vertical, and diagonal directions from fixation. We can see that there are no strong anisotropies and that performance is not systematically worse along the diagonals.

fixate next, this model samples a saccade at random (from empirical data) conditioned on the current fixation location, $S(x', y', x_i, y_i) = p((x', y') \mid (x_i, y_i))$. We base S on the distribution of saccades recorded during the visual-search experiment. To estimate S, we start by quantizing the fixations to a $Q \times Q$ grid ($Q = 32$). This gives a mean of 27 saccades starting from any given position; due to the central bias, this distribution is skewed, with most of the saccades starting in the central cells. 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 it, these saccades are not included in this distribution. In order to deal with the sparsity of the data, we convolve S with a four-dimensional Gaussian filter ($\sigma = 3$). Figure 7 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 IOR or saccadic momentum. Furthermore, the stochastic searcher does not adjust its behavior

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

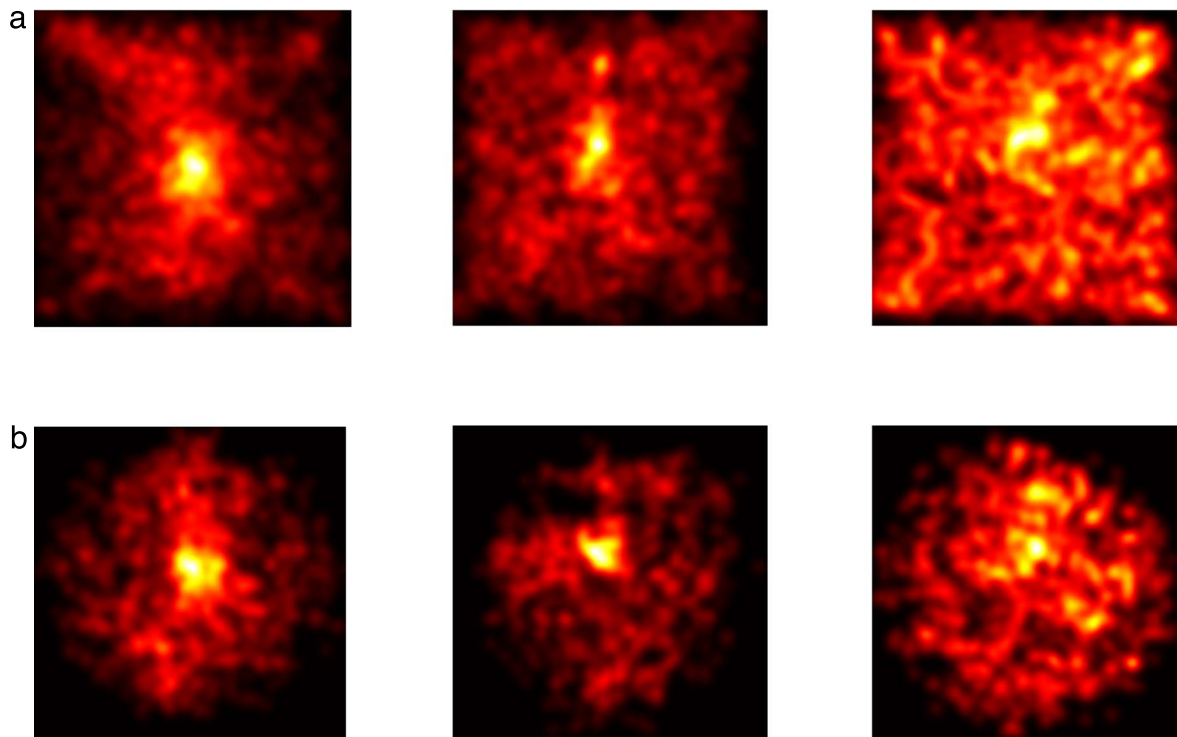Table 1. Coefficients used in Equation 1, the target-detection function.

Figure 5. Hot-spot maps for (a) square and (b) circular search areas for Participants 2, 4, and 5 (left to right). While there are some individual differences—in particular, how strong the central bias is for each observer—there is no clear tendency to fixate above and below the fixation.

based on the difficulty of the search task or the probability of detecting the target.

Comparison between human observers and the model

We now compare the model to human performance by simulating search over the same number as trials used in the human experiment. As Figure 8 shows, the stochastic searcher requires a similar number of fixations to find the target as our human participants.

We can analyze how the number of fixations differs between human and stochastic observers using a general linear mixed-effect model (family = Poisson), with random slopes for β and random intercept: $n_{\text{fix}} \sim \beta * s + (1 + \beta i \mid p_h)$, where s is a two-level factor coding human or model and p_h is the human observer's ID. The only statistically significant difference between the human observers and the stochastic model was for the smooth surfaces, for which the stochastic model required on average half a fixation more than the human observers to find the target.

Figure 9 shows the distribution of saccade amplitudes and directions. Interestingly, we see that even
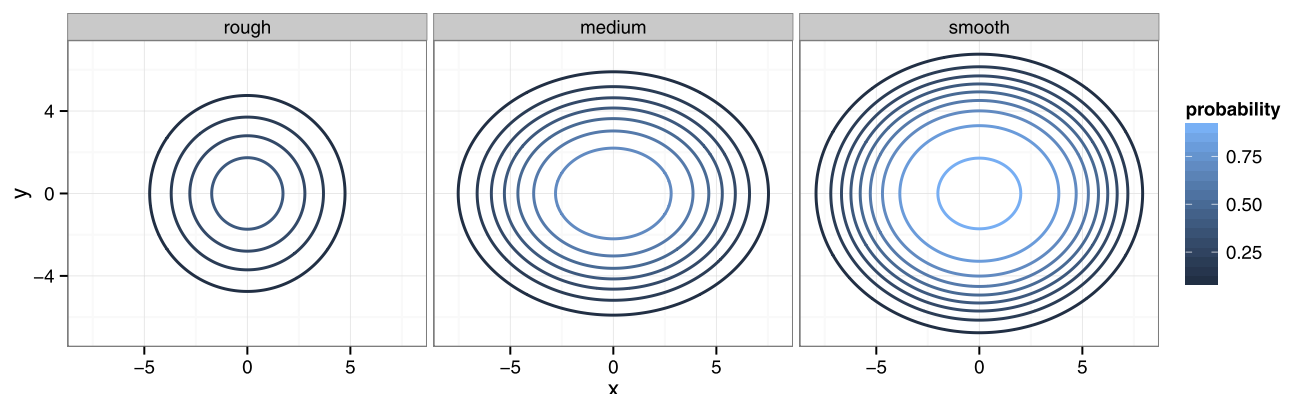

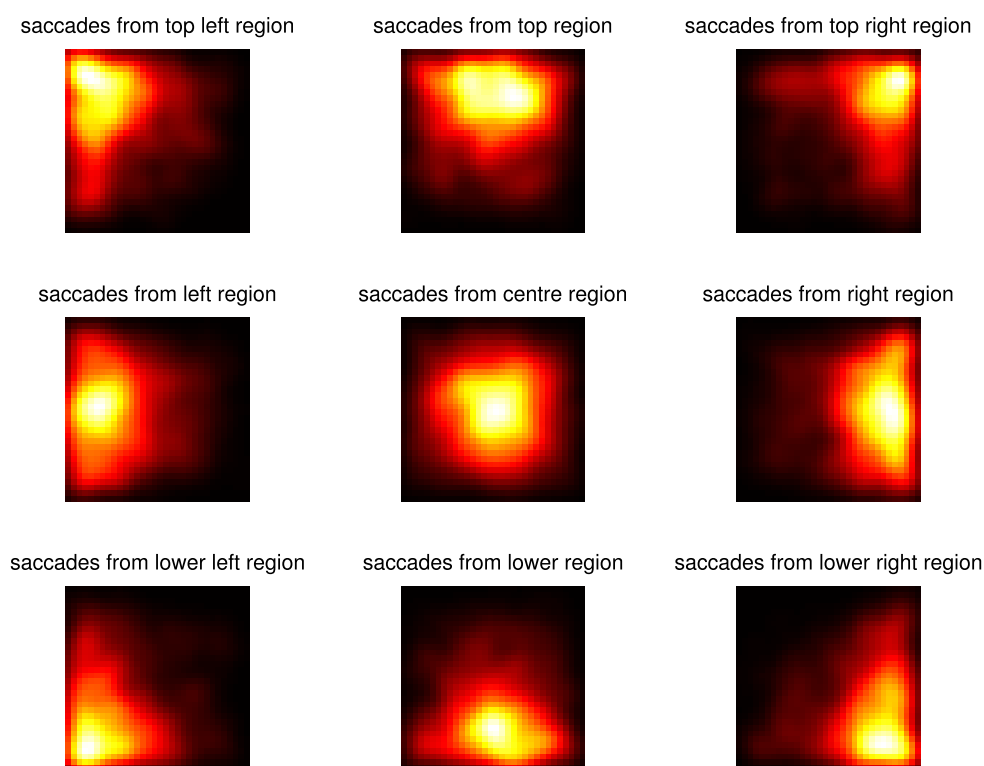Figure 6. Contour plot showing the target-detection model.

Figure 7. Each subplot shows a hot-spot map of fixation locations from a different region of the stimuli. For example, saccades originating from the corner regions tend to be directed back towards the center or along one of the edges.

though the stochastic model is only constrained by the distribution of saccades made by human observers, this is sufficient to give a reasonable match for the displacement over two saccades, and hence it exhibits a similar level of IOR as the human observers. The largest difference between the model and the observers can be seen in the relative direction between two successive saccades. Human observers show a large peak around $0° = 360°$, which indicates saccadic momentum: A second saccade is likely to be made in the same direction as the first saccade.

Discussion

In general, the stochastic model's results provide a good match for human behavior, in terms of the number of fixations to find the target. This conclusion is somewhat surprising given that Najemnik and Geisler (2008) demonstrated that the number of fixations required by observers to find a target is also consistent with an optimal strategy. Najemnik and Geisler also pointed to the distribution of fixation locations as evidence that humans employ an optimal strategy: The model predicts a distribution with peaks above and below the central fixation cross, which they also observed in their three human searchers. However, using a larger sample of partic-

ipants, we found that this result did not replicate. One plausible reason our results from nine observers did not produce this pattern is that our experiments used a square search array, whereas 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, and found that this made no difference.

Reanalysis of Clarke, Green, and Chantler (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 data set (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 data set the target's eccentricity is systematically varied, allowing us to compare the stochastic model to human performance separately for different target eccentricities. As in Experiment 1, we first generated a target-visibility function using a
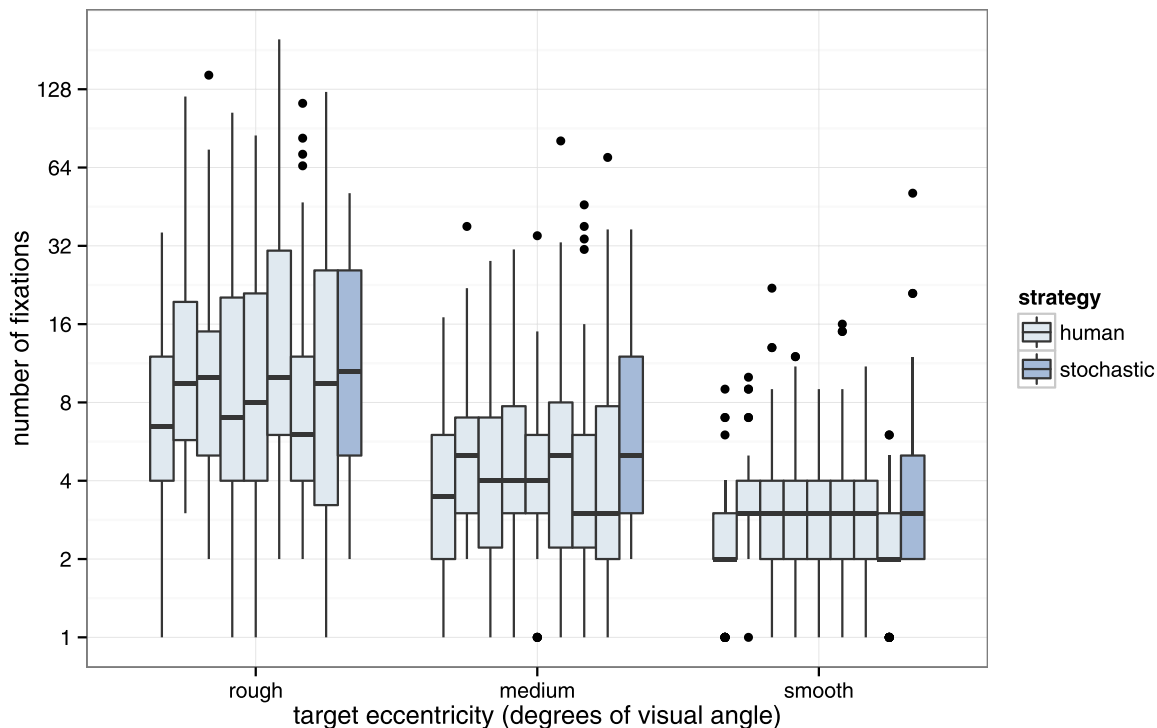
Figure 8. Number of saccades made by human observers and the stochastic search simulation.

target-detection task. We collected target-visibility data from two naïve observers who did not perform the visual-search task.

Target-detection experiment

Observers

Two observers carried out all the trials, split into 20 blocks of 132 trials each, over a number of days. They were paid £50 each. The research was conducted in accord with the Code of Ethics of the World Medical Association (Declaration of Helsinki), and informed consent was obtained from both observers.

Stimuli

Surface textures were created as detailed in Experiment 1. 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 2,160 target-present trials and 480 target-absent trials. This ratio of target-present to target-absent trials ensured that observers made roughly equal numbers of "present" and "absent" responses.

Setup

Observers were instructed to keep their eyes fixated on the center 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 (500 ms), stimulus (200 ms), white-noise mask (500 ms), and finally another fixation cross displayed until a target-present or target-absent response was given. Trials were presented in a random order.

A Tobii x50 eye tracker was used to sample the observers' gaze every 20 ms, and trials were included in the further analysis only if the mean gaze location was within 1° of the central fixation cross and the standard deviation of the gaze's x and y components was less than 0.67°.

Results

A total of 13.6% of trials were removed from analysis (due to breaking central fixation). We first checked the false-positive rates on target-absent trials, which were similar to those seen in Experiment 1 (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 Experiment 1. Results and model coefficients are similar.
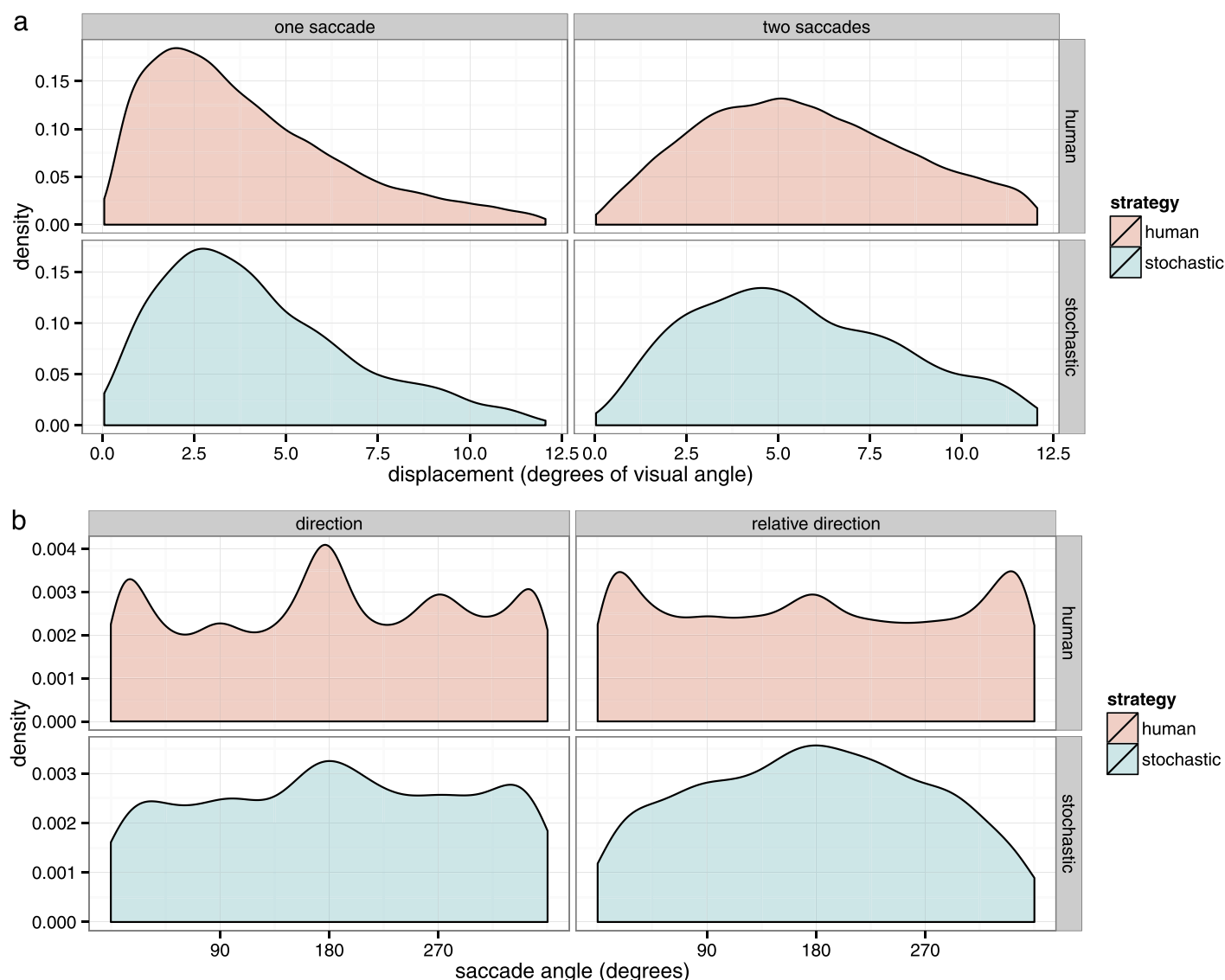
Figure 9. (a) Displacement over one and two saccades. (b) Absolute and relative direction of saccades. The fact that the stochastic model matches the distribution of displacement and direction over a single saccade is a direct result of sampling from empirical data and is therefore unsurprising.

Visual-search data set

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

Observers

Seven observers, aged 18–30 years, were given several practice trials and 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).

Surface stimuli

These stimuli were created as in Experiment 1. For each trial a target was positioned randomly on a circle, centered on the middle of the image, with a radius of $1.7° \pm 0.7°$, $3.8° \pm 0.7°$, or $5.9° \pm 0.7°$ of visual angle.

Setup

Stimulus presentation was controlled by Clearview (Tobii Technology Inc., Stockholm, Sweden). All stimuli were 1024×1024 pixels in size and displayed on
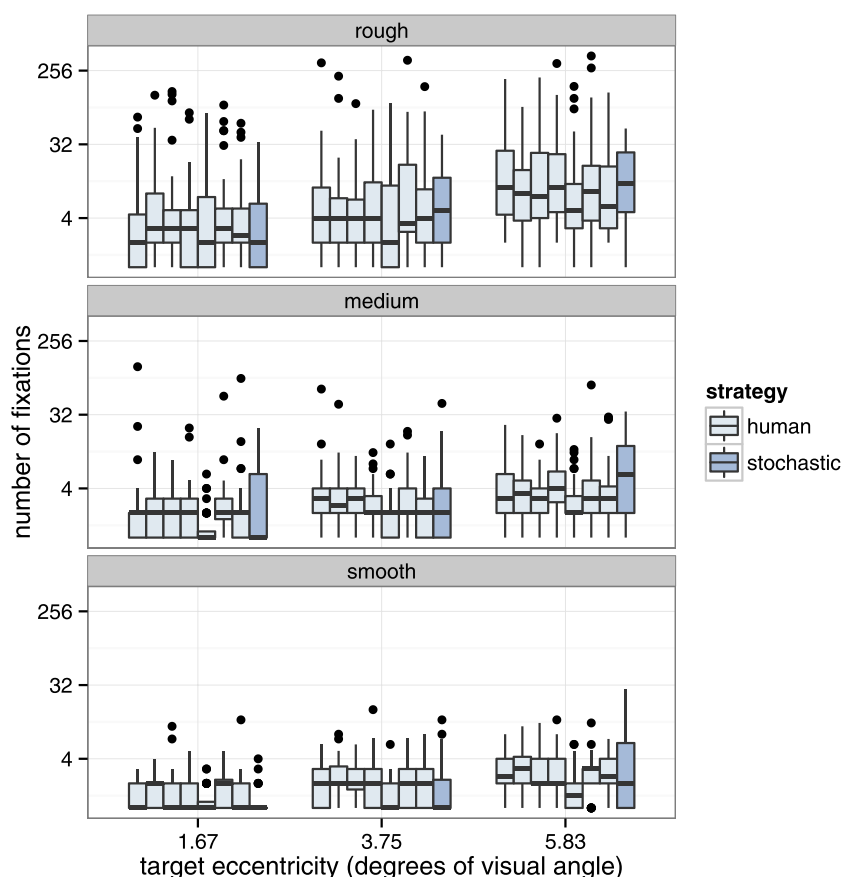
Figure 10. Number of saccades made by human observers and search simulations. The stochastic search model is sufficient to explain the number of fixations required to find the target.

an NEC LCD2090UXi monitor. The pixel dimensions were 0.255 × 0.255 mm, resulting in images with physical dimensions of 26.1 × 26.1 cm. The monitor was linearly calibrated with a GretagMacBeth Eye-One; maximum luminance was set at 120 cd/m². 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 87 cm away from the display monitor. At this distance, 1 pixel is approximately 1 arcmin of visual angle; images subtend 16.7° of visual angle, and the targets 0.66°.

Results

The number of fixations required to find the target is shown in Figure 10. Using a generalized 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.

Comparison with search models

We now compare the model to human performance by simulating search over the same number as trials as used in the original experiment (incorrect and target-absent trials are discarded). The results closely match those of Experiment 1. In terms of the number of saccades required to find the target, we find that the stochastic searcher offers very similar performance to the seven human observers over the range of surface roughnesses and target eccentricities used (Figure 10).

General discussion

Human visual-search performance can be modeled by a stochastic process. Beyond the number of fixations required to find the target, the stochastic model also does a reasonable job of explaining the spatial distribution of fixations, the frequency of saccade

amplitudes, and the relative angles of sequences of saccades.

It is interesting that the stochastic and optimal models—two models that have very different architectures and produce very different search behavior—could both take a similar number of fixations as humans to detect a target. Clearly there is more than one way to achieve this single end, and maximizing 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 did either the optimal model or human searchers. However, their random model selected uniformly 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 not take into account the natural tendencies in saccade behavior 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.

An important difference between our study and those of Najemnik and Geisler (2005, 2008) is in the experiment used to collect data to create the target-visibility map. Najemnik and Geisler 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. Cuing a target's location has long been known to improve performance (Posner, 1980), possibly by enhancing the target signal (Yeshurun & Carrasco, 1998). Indeed, our pilot experiment comparing these two measures of sensitivity demonstrated that the 2AFC method inflates sensitivity, particularly in the difficult search conditions, and reduces differences between difficulty manipulations. 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. Another key difference is that Najemnik and Geisler (2008) report a tendency to fixate above and below the center of the search display for both human observers and their optimal model. This tendency did not appear in our human data, even when we reran the search experiment using a circular search display to match the one used by Najemnik and Geisler (2008).

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 behavior. 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 is 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 IOR to achieve humanlike behavior, we do not mean to suggest there is no IOR 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 the search array's boundaries, rather than being applied to discrete search objects. IOR is strengthened by objects (Jordan & Tipper, 1998), and this may be particularly true when the eyes move, necessitating inhibition of spatiotopic, rather than retinal, coordinates (e.g., Krüger & Hunt, 2013). We echo the sentiments of Najemnik and Geisler (2005) in their conclusions in noting that our stochastic model complements, rather than contends with, existing models of search. While our model provides a good match for search in this limited context, an important question for future work is how a stochastic saccade-selection process combines with other search mechanisms, such as salience, IOR, and contextual cuing, in more complex search situations (e.g., Farrell et al., 2010).

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 saccades 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. As far as we can tell from our data, the difficulty of the current search task only has a limited effect on the distribution of saccades (median saccadic amplitude increases from 3.2° for the easy trials to 3.9° for the most difficult condition, but there is considerable overlap between distributions). The relative similarity of the saccade amplitudes across search difficulty suggests that, at least in the context of our search arrays, the set of possible saccades and selection from this set is not very sensitive to the search context. The mechanism that has been assumed to subserve efficient eye-movement behavior in natural tasks is reinforcement learning (e.g., Hayhoe & 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's 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 that humans are suboptimal in search (Morvan & Maloney, 2012; Verghese, 2012; Zhang et al., 2012), in that saccades during search do not appear to maximize the probability of detecting the target based on previous fixations and knowledge about the limits of our own visual acuity. Fortunately, our results suggest that this kind of optimality is not a prerequisite for search efficiency: A stochastic model can perform as well as human observers.

Keywords: visual search, saccades, optimal behavior

Acknowledgments

Commercial relationships: none.
Corresponding author: Alasdair D. F. Clarke.
Email: a.clarke@abdn.ac.uk.
Address: School of Psychology, University of Aberdeen, Aberdeen, UK.

Footnote

[1] Participant 1 did 1,799 trials, and Participant 7 did 2,439 trials. These were the first two participants to carry out this part of the experiment, and adjustments to the number of trials were made to make the experiment last approximately 1.5 hr.

References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. In: *R package version, 1*(7).

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433–436.

Clarke, A. D. F., Green, P. R., Chantler, M. J., & Emrith, K. (2008). Visual search for a target against a $1/f^{\beta}$ continuous textured background. *Vision Research, 48*(21), 2193–2203.

Clarke, A. D. F., Green, P. R., & Chantler, M. (2009). Modeling visual search on a rough surface. *Journal of Vision, 9*(4):11, 1–12, doi:10.1167/9.4.11. [PubMed] [Article]

Clarke, A. D., & Hunt, A. R. (2016). Failure of

intuition when choosing whether to invest in a single goal or split resources between two goals. *Psychological Science, 27*, 67–74. doi:10.1177/0956797615611933.

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, doi:10.3758/BF03195489.

Farrell, S., Ludwig, C. J., Ellis, L. A., & Gilchrist, I. D. (2010). Influence of environmental statistics on inhibition of saccadic return. *Proceedings of the National Academy of Sciences, 107*(2), 929–934.

Fox, J., & Weisberg, H. S. (2010). *An R companion to applied regression*. London: Sage Publications.

Gao, D., Mahadevan, V., & Vasconcelos, N. (2008). On the plausibility for the discriminant center-surround hypothesis for visual saliency. *Journal of Vision, 8*(7):13, 1–18, doi:10.1167/8.7.13. [PubMed] [Article]

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):25, 1–18, doi:10.1167/9.5.25. [PubMed] [Article]

Itti, L., & Baldi, P. F. (2009). Bayesian surprise attracts human attention. *Vision Research, 49*(10), 1295–1306.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10), 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), 735–744.

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 J. Braun, C. Koch, & J. Davis (Eds.), *Visual attention and neural circuits* (pp. 159–175). Cambridge, MA: MIT Press.

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

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

Over, E. A. B., Hooge, I. T. C., Vlaskamp, B. N. S., & Erkelens, C. J. (2007). Coarse-to-fine eye movements strategy in visual search. *Vision Research, 47*(17), 2272–2280.

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*(9): 429, doi:10.1167/3.9.429. [Abstract]

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

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, R. P., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Research, 42*(11), 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):6, 1–20, doi:10.1167/7.5.6. [PubMed] [Article]

Stone, M. (1960). Models for choice-reation time. *Psychometrika, 25*(3), 251–260.

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision, 7*(14):4, 1–17, doi: 10.1167/7.14.4. [PubMed] [Article]

Tatler, B. W., & Vincent, B. T. (2009). The prominence of behavioural biases in eye guidance. *Visual Cognition, 17*(6–7), 1029–1054.

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 W. Gray (Ed.), *Integrated models of cognitive systems* (pp. 99–110). New York: Oxford University Press.

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.

Appendix

Yes/no versus 2AFC

Here, we compare and contrast the visibility maps obtained via the target-detection task used in the study with 2AFC, similar to that used by Najemnik and Geisler (2005, 2008). We used one naïve observer, who carried out the 2AFC task first and then the target-detection task. This means that any practice effects will mainly inflate the target-detection experiment.

Methods

All stimuli were the same as in the main experiments.

2AFC

Following Najemnik and Geisler (2005, 2008), the experiment was run in a number of blocks. Within each block, the target was presented at a single known location. The observer was shown two stimuli, each for 250 ms, with an interstimulus interval of 750 ms. The observer was then required to respond using the keyboard and indicate whether the target had been present in the first or second stimuli. The experiment was run over several sessions, six blocks at a time. Only target positions along the horizon were tested.

Yes/no

This part of the experiment was identical to the target-detection task presented in the main text.

Results

The results shown in Figure A1 show a clear difference between methods: The d' values obtained from the 2AFC task are largely insensitive to background difficulty. This intuitively makes sense. In the 2AFC task the target's location is known to the observer, allowing the deployment of covert attention to the correct location before stimulus onset.

We can confirm this difference with a two-way ANOVA. We find that for the 2AFC data, there is a statistically significant effect of target eccentricity on d', $F(1) = 67.7$, $p < 0.001$, but no significant effect of β, $F(2) = 0.81$, or of the interaction, $F(2) = 0.80$. However, when looking at the target-detection results, we find statistically significant effects of target eccentricity, $F(1) = 70.3$, $p < 0.001$, and β, $F(2) = 54.8$, $p < 0.001$. Again, there was no statistically significant interaction, $F(2) = 0.09$.

Discussion

We find that the two methods for measuring the target's visibility map give qualitatively different results. The d' scores obtained using 2AFC appear to be insensitive to changes in surface texture. This is at odds with the results of the visual-search data presented in this article (Figure 8) and in reports by Clarke et al. (2008; 2009), where we can see that surface roughness has a large effect on search difficulty. In particular, we see that the observer had a much higher d' for rough surfaces under the 2AFC measurement, which would be expected given the literature on covert attention. For smoother surfaces, the opposite effect is observed. This is potentially due to practice effects. Either way, the implications of underestimating target visibility for smooth cases are minimal, as observers are already finding the target with only one or two fixations.
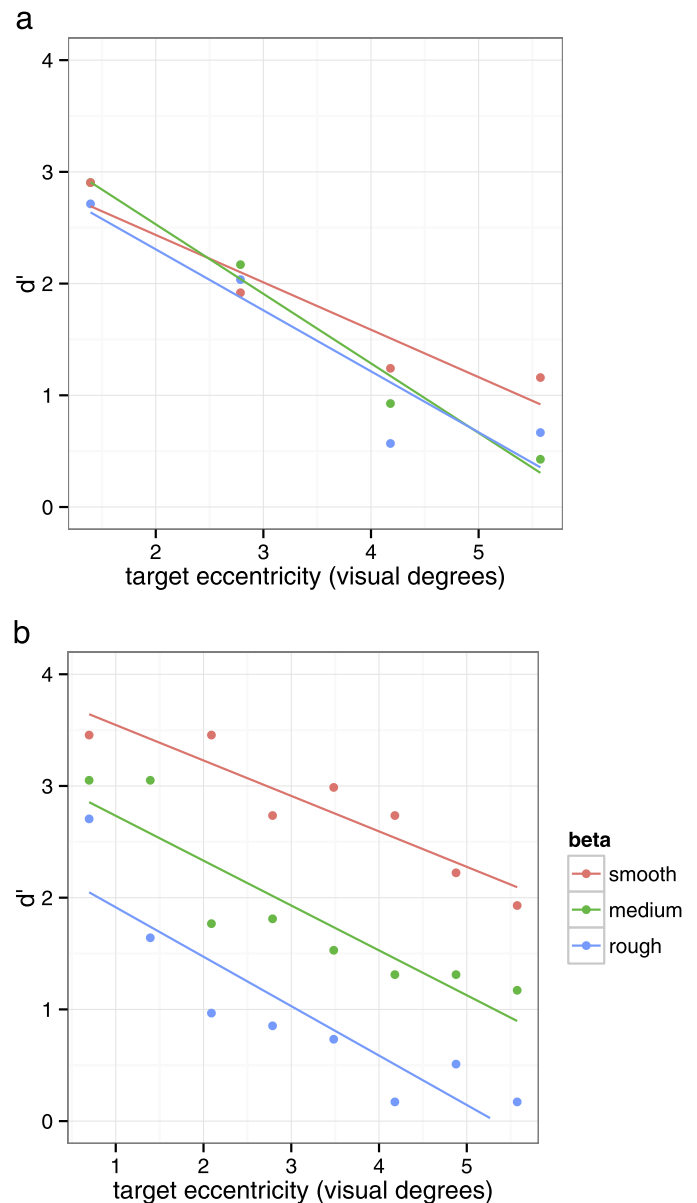
Figure A1. Results from the comparison of 2AFC and yes/no.