A Suboptimal Model Explained the Spatiotemporal Control of Human Eye Movements in Visual Search

Yunhui Zhou[1], Yuguo Yu[1,2*]

[1] School of Life Sciences, Fudan University, Shanghai 200433, P.R. China

[2] Human Phenome Institute, Institutes of Brain Science, Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai 200433, P.R. China.

[*] Correspondence should be addressed to:

Yuguo Yu, Ph.D.

Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai 200433, China. Email: yuyuguo@fudan.edu.cn

Author contributions: Y.Y. supervised the research, Y.Z. and Y.Y. designed the research, Y.Z. performed the research, wrote the analysis tools and analysed the data, Y.Z. and Y.Y. wrote the paper.

Conflict of Interest: The authors declare no competing financial interests.

Acknowledgments: This work was supported by the National Natural Science Foundation of China (81761128011), the Shanghai Municipal Science and Technology 795 Major Project (No. 2018SHZDZX01) and ZJLab, the program for the Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning. We thank Professor Zhaoping Li for her advice on the manuscript.

Abstract

Efficient eye movement planning in both space and time is vital for visual perception. Previous studies showed that human eye movements were consistent with an optimal searcher during visual search; however, we found that such a model could not fully explain the spatial control of eye movements and it completely ignored the temporal control aspect. Here, we first measured the temporal course of human visibility map, and then measured eye movement dynamics during a visual search task. We further built a continuous-time eye movement model that determined both fixation duration and location. When the participants' saccade amplitude bias and memory capacity limitations were introduced, this model could describe multiple aspects of human visual search behavior much better than the optimal searcher with fixed fixation duration. Interestingly, this model predicted that the human visual memory capacity is approximately 8 fixations. The results showed that human performed suboptimal in visual search, and implied that human may use a mixed control strategy to select fixation location. This strategy may have adaptive advantages considering the limitations of the human brain and statistics of the natural environment.

Introduction

For animals with foveated retina, efficient visual search is important for survival, and requires close-to-optimal planning of eye movement both in space and time. Currently, there are conflicting evidence on whether human can make spatially optimal eye movements. Najemnik & Geisler (2005, 2008, 2009) have shown that humans can perform close to optimal models in a multi-saccade visual search task, but other studies indicated that human may rely more on suboptimal heuristics rather than calculating and planning optimal eye movement for a specific task (Araujo et al., 2001; Morvan and Maloney, 2012; Nowakowska et al., 2017). Optimally planning eye movements has the advantage of analyzing the environment at a faster speed, but following simple heuristics also has the advantage of conserving computational power, so in reality human may trade off in between. Besides, the ideal searchers proposed by (Najemnik and Geisler, 2005, 2009) have unlimited visual memory, so it is impossible that human strictly follows the optimal search rule. We therefore reconsider the question of how optimal do humans search, and ask if a visual search model with more biological limitations can perform more closely to humans.

Moreover, there is a surprising lack of consideration of time in visual search models. For example, the Target Acquisition Model does not explain fixation duration (Zelinsky, 2008, 2012); the ideal searchers assume each fixation last for 250 ms (Najemnik and Geisler, 2005, 2009); and the Guided Search model assumes fixation duration is 200 – 250 ms (Wolfe and Gancarz, 1997). We found only one model that considered fixation times (Halverson and Hornof, 2011), but the model did not explicitly model how information accumulates over time, and only explained the mean fixation duration. However, for visual tasks like reading and free viewing, there are models that fully capture the fixation duration distribution by

simulating the information accumulation process (Engbert et al., 2005; Nuthmann et al., 2010; Trukenbrod and Engbert, 2014; Tatler et al., 2017), but the temporal accumulation of visual information in these models are usually artificial rather than based on experiment data. What's more, eye movement models of scene viewing do not actually acquire information from the stimulus, instead they simply fit eye movement metrics *per se* by image features, so it is unknown whether these models can actually perform visual search with human-level performance. To bridge the gap, we believe that a model of visual search that fully explains the spatiotemporal control of eye movements should combine the merits from existing models of both visual search and scene viewing.

In this research, we experimentally measured the accumulation of visual information by the brief stimulus exposure method. The data was then used to formulate a drift-diffusion process that decides saccade timing. Combined with the ideal search theory, we proposed a continuous-time optimal searcher that determines both fixation location and fixation duration. We found that human participants do not follow the optimal search strategy. Then, we derived a suboptimal model with a constraint on the saccade amplitude and memory capacity that better agrees with participants' eye movement statistics compared to the continuous-time ideal searcher.

Materials and Methods

Participants

Seven student volunteers (4 males and 3 females, age range 19-26 years) participated in the experiment for monetary compensation. One subject was the author (Y.Z). All participants had normal or corrected-to-normal vision. This study was approved and

supervised by the Ethics Committee of the School of Life Sciences at Fudan University (No. 290), and written informed consent was obtained from all participants.

Apparatus

Stimuli were presented on a 24.5-inch BenQ ZOWIE XL2540 LCD monitor at a resolution of 1920×1080 and a refresh rate of 240 Hz. Participants were seated 70 cm in front of the display with their heads fixed by a forehead and chin rest. Stimuli were generated and presented using MATLAB R2017b (MathWorks) and the Psychophysics Toolbox extensions version 3.0.15 (Brainard, 1997; Pelli, 1997; Kleiner M, Brainard D, 2007) running on a Windows 7 (Microsoft) machine. Eye movements of both eyes were recorded at 2000 Hz with a TRACKPixx3 eye-tracker (Vpixx Technologies, Canada).

Experiment Design and Statistical Analysis

Target and background image. The target was a six cycle-per-degree Gabor with a diameter of 0.3 degrees, oriented 45 degrees counterclockwise from the vertical, and windowed by a symmetrical raised cosine (half-height width was one cycle of the Gabor). The background image was a circular naturalistic noise image (whose statistics follow a $1/f$ power spectrum) placed at the center of the screen with a diameter of 15 degrees and root-mean-squared (RMS) contrast of 0.2 (Figure 1A). The RMS contrast is defined as the standard deviation of the pixel gray value divided by the mean gray value. The area outside the background image was set to the mean gray value of the image (0.5). The background images used in each trial of detection task and visual search task were randomly picked from a database of 1000 independently generated images.

Eye-tracker calibration. During each laboratory visit, participants first performed a 13-point calibration routine. The outer edge of the 13 points formed a square that covered the

circular search area. We repeated the calibration until the average test-retest measurement error of both eyes fell below 0.5°. Recalibration of the eye-tracker was performed whenever the eye-tracking accuracy failed to meet the requirements of the experiment (described below).

Detection task. The detection experiment contained two sections. In the first section, we measured the relationship between foveal target visibility and target RMS contrast and selected the target contrast level that made the foveal visibility equal to 3. Each participant therefore had a unique target contrast and was used for all subsequent experiments. The goal was to normalize the target detection difficulty across participants. Before the start of each trial, a black cross and stroked circle appeared at the center of the screen, indicating the fixation location and target location. The remaining display area was set to a 0.5 gray value. Participants began a trial by fixating within 1° from the screen center and pressing a button. The cross and circle then disappeared for 200−400 ms (chosen randomly), followed by two 250 ms intervals of the stimulus image separated by a 500 ms gray interval. The two stimulus intervals had the same noise background image; however, a randomly chosen stimulus had a target at the screen center. The participants then chose the interval that contained the target with a keyboard. A trial was aborted if the participant fixated more than 1° away from the screen center at any time during the trial. This section contained 4−5 levels of target contrast (depending on the participant's pace), and each level included 200 trials. Trials with different target contrasts were randomly interleaved. We chose the appropriate target contrast levels (ranging from 0.03 to 0.12) by a pilot experiment for each participant to avoid floor or ceiling effects.

In the second section, we measured the relationship between the stimulus presentation

time and target visibility along four cardinal directions in the search field. For each participant, the target RMS contrast was the value obtained from section one. There were two versions of experiment in this section. Four participants participated in the first version, four participated in the second version, with one participant (the author Y.Z) participated in both. The procedure was the same as in section one except that we varied both the stimulus duration and target location. In the first version of experiment, participants were required to fixate at the screen center and decide which of the two stimulus intervals contained the target at the cued location. The experiment contained five blocks (four cardinal directions plus the screen center). For each block, we chose 3−4 different target locations (except when the target was at the screen center) and 4−5 levels of stimulus presentation time. The target was located 0−6.75° (0−90% of the image radius) away from the screen center. In the second version of experiment, the target location cue was not shown, and in each trial, target location was randomly chosen from all possible locations in the first version of experiment. When the target appeared at the screen center, the stimulus duration was chosen from 4 to 200 ms, and each level was tested for 100 trials. When the target appeared at peripheral locations, the stimulus duration was chosen from 50 to 700 ms, and each combination of stimulus length and target location was tested for 50 trials. In the main text we report data from the four participants in first version of the experiment. Data from the second version was reported in supplemental material.

The entire detection task contained approximately 5000 trials. Participants finished these trials by multiple laboratory visits within two months. Before the start of the formal experiment, participants practiced with approximately 100 trials to stabilize their performance. Participants received feedback about their performance during practice trials

but did not receive feedback during the formal experiment.

Visual search task. In this experiment, participants were asked to search for the target within the background noise image as fast as possible while trying to maximize the response accuracy. The target locations were sampled uniformly inside the search area, and participants were informed about this feature. Participants began each trial by fixating within 1° of the screen center and pressing a button. Then, the search images appeared and waited for participants' responses. There was no upper limit on the waiting time. When the participants found the target, they fixated at the target and pressed a key on the keyboard. The final fixation location and the target's true location were then shown on the screen. If the distance between these two locations was less than 1°, we labeled the response as correct. Participants were asked to report if they had found the correct target location but the response was judged as incorrect. In this case we would recalibrate the eye tracker. Participants performed 200−330 (mean = 235) visual search trials depending on their pace and available time.

Selecting target RMS contrast. We first calculated the hit rate and correct rejection rate of each target contrast level from behavior data in section one of the detection tasks. The hit (or correct rejection) rate was defined as the number of trials that target appeared in the first (or second) interval, and the participant made a correct response, divided by the total number of trials that target appeared in the first (or second) interval. The relationships between the hit/correct rejection rate and target RMS contrast c were then separately fit with a Weibull function:

$$f(c) = 0.5 + 0.5 \times \left\{ 1 - \exp\left[-\left(\frac{c}{c_T} \right)^s \right] \right\} \tag{1}$$

The threshold parameter c_T and steepness parameter s were estimated by maximum-likelihood methods implemented in the Palamedes Toolbox (Prins and Kingdom, 2018). The relationship between the foveal target visibility and target RMS contrast could then be estimated by signal detection theory (Macmillan and Creelman, 2005):

$$d'(c) = \frac{z\left[f_H(c)\right] - z\left[1 - f_{CR}(c)\right]}{\sqrt{2}} \tag{2}$$

Here, $z(x)$ is the inverse standard normal cumulative distribution function. $f_H(c)$ and $f_{CR}(c)$ are the estimated hit and correct rejection rates given by Equation 1. We could then obtain the target contrast level that made $d'(c) = 3$.

Fitting the temporal course of the visibility map. Behavior data from different participants in the second section of the detection task were first pooled together. Target visibility at each location for each level of stimulus duration was then calculated by Equation 2. The relationship between the visibility d', target location (x,y) relative to the fixation location, and stimulus duration t were then fit by the following equations:

$$d'(x,y,t) = a(x,y) \cdot \sqrt{\frac{1 - \exp\left[-k(x,y) \cdot t\right]}{k(x,y) \cdot \left\{1 + \exp\left[-k(x,y) \cdot t\right]\right\}}} \tag{3}$$

$$a(x,y) = p_1 \cdot \exp\left(-p_2 \cdot \sqrt{x^2 + p_5 \cdot y^2}\right) \tag{4}$$

$$k(x,y) = p_3 \cdot \exp\left(-p_4 \cdot \sqrt{x^2 + p_5 \cdot y^2}\right) \tag{5}$$

The parameters p_1, p_2, p_3, p_4, p_5 in these equations were estimated by minimizing the MSE between the calculated and measured visibility data using the GlobalSearch algorithm in the Global Optimization Toolbox of MATLAB. The derivation of these equations is shown in the supplemental material.

Eye-tracking data analysis. Eye-tracking data were analyzed by the EYE-EEG

extension (Dimigen et al., 2011) of the EEGLAB toolbox (Delorme and Makeig, 2004). Only the data from the dominate eye (determined by a hole-in-the-card test) were analyzed. All participants were right-eye dominant. We used the algorithm described by Nyström & Holmqvist (2010) with default parameters to classify eye movements into saccades and fixations. In the main text we report data from the four participants in first version of detection task. Across these participants, the algorithm identified 15,201 fixations and 16,269 saccades. However, we visually inspected the data and excluded 12 fixations (.08%) and 15 saccades (.09%) that originated from recording noise. The data from another four participants in the second version of detection task was reported in supplemental material. The algorithm by Nyström & Holmqvist excludes fixations interrupted by blinks or noise; therefore, the number of detected fixation events was smaller than the number of detected saccade events.

When summarizing fixation metrics, we did not include the last fixations (when a response was made) for fixation duration to avoid the effects of the participants' motion on the fixation duration. When collecting fixation locations, we did not include the initial fixations because, in all the trials, the search started from the screen center. We did not further remove any saccades when summarizing the saccade metrics.

The Entropy Limit Minimization (ELM) Searcher

We used the ELM searcher (Najemnik and Geisler, 2009) as our baseline model. This searcher selects a fixation location to maximize the expected information gain of the next fixation (equation 11 below). The ELM searcher greatly reduces the computational burden compared to the Bayesian ideal searcher while maintaining a similar search performance (Najemnik and Geisler, 2005, 2009) and has received electrophysiological

support (Horan et al., 2019). The model can only fixate at 400 locations uniformly sampled inside the search field (Figure 2B). We set the fixation duration to 250 ms and evaluated the visibility map using Equations 3-5. The model terminates the search whenever the posterior probability of the target being at the current fixation location exceeds a threshold TH_{targ}. The value of TH_{targ} was chosen to make the model's correct response rate comparable to the participants' data.

The Continuous-time ELM (CTELM) Searcher

We propose the CTELM searcher, which extends the ELM searcher to account for the distribution of fixation durations. A schematic diagram of the CTELM searcher is shown in Figure 2A and will be described in the following sections.

Temporal Course of Evidence Accumulation. The accumulation of visual evidence over time should be derived according to the change in target visibility over time. In brief, we formulate a drift-diffusion process W_t at each location with the following properties:

(1) At any time t, W_t follows a normal distribution.

(2) At any time t, the expectation and standard deviation obeys $\mathrm{E}[W_t] = -\mathrm{E}[W_t^{'}]$ and $\mathrm{SD}(W_t) = \mathrm{SD}(W_t^{'})$, where W_t is generated by target plus noise, and $W_t^{'}$ is generated by only noise at this location.

(3) At any time t, $2\mathrm{E}[W_t]/\mathrm{SD}(W_t)$ equals the target visibility $d^{'}$ measured at this location. This is the definition of $d^{'}$ in signal detection theory (Macmillan and Creelman, 2005).

In practice, we calculate W_t as the leaky integration of independent normal random variables ΔW generated within each time interval Δt. If the target appears at this location, ΔW follows the $N(\Delta t, \sigma^2\Delta t)$ distribution; otherwise, $N(-\Delta t, \sigma^2\Delta t)$. The visibility of the target

at this location is then controlled by the parameter σ.

In the supplemental material, we show that if ΔW are integrated perfectly over time, the visibility at any location will approach infinity given an infinite viewing time. This is impossible in reality. Therefore, we assume that the weight w of previous evidence decays exponentially with a speed parameter k over time:

$$w(t) = \exp(-k \cdot t) \tag{6}$$

The total accumulated evidence W_i at the ith timestep can then be calculated as

$$\begin{cases} W_0 = 0 \\ W_i = \Delta W + W_{i-1} \cdot w(\Delta t) \end{cases} \tag{7}$$

In the supplemental material, we show that the values of the parameters k and σ depend on the queried position (x,y) relative to the fixation location. The value of k is calculated by Equation 5. The parameter σ is calculated by $\sigma = 4/\sqrt{2}a$, and the value of a is calculated by Equation 4.

Calculating Target's Posterior Probability Map. At each time step, the posterior probability of target being at the ith ($i = 1 \cdots n$, $n = 400$) location can be calculated from the accumulated evidence from the current and previous fixations by Bayesian theory:

$$P_{i,T_F} = P\left(i \mid \mathbf{W}_{L_1,T_1}, \cdots, \mathbf{W}_{L_F,T_F}\right) = \frac{P\left(\mathbf{W}_{L_1,T_1}, \cdots, \mathbf{W}_{L_F,T_F} \mid i\right) p(i)}{\sum_{j=1}^{n}\left[P\left(\mathbf{W}_{L_1,T_1}, \cdots, \mathbf{W}_{L_F,T_F} \mid j\right) p(j)\right]} \tag{8}$$

Here, the fixations are indexed by $1 \cdots F$, and L_F and T_F represent the location and duration of the F_{th} fixation, respectively. $p(i) = 1/n$ is the prior probability of the target being at location i. The vector $\mathbf{W}_{L_F,T_F} = \left(W_{1,L_F,T_F}, \cdots, W_{n,L_F,T_F}\right)$ is the accumulated evidence by time T_F of the F_{th} fixation at all locations. Equation 8 implies that the CTELM searcher has perfect memory

across fixations.

Assuming that visual information is independently accumulated at each location during each fixation, Equation 8 can be simplified to (see the supplemental material for details)

$$P_{i,T_F} = \frac{p(i)\cdot\exp\left[\sum_{f=1}^{F}\frac{a_{i,L_f}^2\cdot W_{i,L_f,T_f}}{2+2\exp\left(-T_f\cdot k_{i,L_f}\right)}\right]}{\sum_{j=1}^{n}\left\{p(j)\cdot\exp\left[\sum_{f=1}^{F}\frac{a_{j,L_f}^2\cdot W_{j,L_f,T_f}}{2+2\exp\left(-T_f\cdot k_{j,L_f}\right)}\right]\right\}} \tag{9}$$

Here, a_{i,L_f} and k_{i,L_f} refer to the parameters *a* and *k* in Equations 4 and 5 at location *i* during fixation *f*. By applying Equation 9 to all target locations at each time step, we can calculate the change in the posterior probability map of the target location over time.

Decision on Saccade Timing. In the model, saccades are triggered randomly in two different ways according to their relative frequency. The majority (97%) of saccades (normal saccades) are triggered by a decision process concerning whether the current fixation location contains the target. The first saccade in a trial is always a normal saccade. If the probability of the target being at the current fixation location is lower than a threshold, the model makes a saccade decision. We implemented the saccade threshold $TH_{sac}(t)$ as a "collapsing bound", similar to certain decision-making models (Gluth et al., 2012; Ratcliff and Frank, 2012; Thura et al., 2012; Zhang et al., 2014):

$$\log_{10}\left[TH_{sac}(t)\right] = q_1\cdot\exp\left[-\left(\frac{t}{q_2}\right)^{q_3}\right] \tag{10}$$

Here, q_1, q_2, q_3 (q_1, $q_3 < 0$, $q_2 > 0$) are parameters whose values are estimated by fitting to distributions of fixation durations from experimental data. If the next saccade is also a

normal saccade, a new saccade decision process at the next fixation location starts immediately after a saccade decision is made. Although the model controls the fixation duration by foveal analysis, saccade timing is not independent from the peripheral analysis. This is because the fixation duration is controlled by the posterior probability, and, at any time, the posterior probability at all locations sums to one. If the target is very likely to present at a peripheral location, this will lower the probability that the target is at the current fixation location and will terminate the current fixation earlier.

In addition to normal saccades, a small group of low-latency saccades can be generated by fast subcortical response through superior colliculus (Schiller et al., 1987). In the model, we set the frequency of low-latency saccades to 3% according to the frequency of express saccades in an overlap task (Munoz et al., 1998). The latency (in seconds) is directly drawn from a gamma distribution with shape parameter 8 and scale parameter 6.9×10^{-4}, plus a delay of 0.03 seconds. This latency follows a distribution with mean = 85 ms, 5% percentile = 58 ms, and 95% percentile = 121 ms, which roughly falls in the 80-120 ms range of express saccade latency (Fischer and Ramsperger, 1986). If the next saccade is a normal saccade, a new saccade decision process at the next fixation location starts immediately before the delay of 0.03 seconds.

In the model, the delay from the retina to the visual cortex (eye-brain lag) is set to 0.06 seconds (Nowak and Bullier, 1997), and the delay from a normal saccade decision being made to the eye starting to move (saccade lag) is set to 0.03 seconds (Bruce et al., 2017). At the start of a new fixation, the model continues to accumulate information from the previous fixation until the arrival of new information after the eye-brain lag. We do not consider the duration of a saccade itself (a saccade is finished instantly) because of saccadic masking

(Bridgeman et al., 1975).

Decision on Saccade Target. As in the ELM searcher, we select the location that maximizes the expected information gain of the next fixation to be the saccade target:

$$L_{f+1} = \arg\max_{L_{f+1}} \left\{ H_{L_f} - \mathrm{E}\left[H_{L_{f+1}} \right] \right\} = \arg\max_{L_{f+1}} \left[\frac{1}{2} \sum_{i=1}^{n} \left(P_{i,T_f} \cdot d'^{2}_{i,L_{f+1}} \right) \right] \tag{11}$$

Here, H_{L_f} is the entropy of the target location at the current fixation, and $\mathrm{E}\left[H_{L_{f+1}} \right]$ is the expected entropy of the target location if the next fixation location is L_{f+1}.

The deviation of Equation 11 is shown in Najemnik and Geisler (2009), and requires $d'_{i,L_{f+1}}$ to be a fixed value. However, in the model, $d'_{i,L_{f+1}}$ changes with time. As a simplification, we set the expected duration of the next fixation to be 0.25 seconds when evaluating Equation 11. We ran Monte Carlo simulations and found that the expected information gain calculated in this way had a high Pearson correlation (mean = 0.952, min = 0.912, and max = 0.968) with the actual information gain (Figure S4 in the supplemental material).

Decision on Target Location. The search stops whenever the posterior probability at the current fixation location exceeds a target detection threshold TH_{targ}, and the model sets the current fixation location as the target location. The value of TH_{targ} was chosen to make the model's correct response rate comparable to the participants' data.

The Constraint-CTELM searcher

The constraint-CTELM searcher is a suboptimal searcher built to achieve a better fit to the experimental data. The schematic diagram is the same as the CTELM searcher (Figure 2A); however, we constrained its saccade amplitude distribution and the number of previous fixations when calculating the posterior probability of the target location.

Decision on Saccade Target. We used a penalty function $H(L_f, L_{f+1}, f)$ to modify the expected information gain across the search field before selecting the maximum:

$$L_{f+1} = \arg\max_{L_{f+1}} \left[H\left(L_f, L_{f+1}, f\right) \cdot \frac{1}{2} \sum_{i=1}^{n} \left(P_{i,T_f} \cdot d'^{2}_{i,L_{f+1}} \right) \right] \tag{12}$$

$$H\left(L_f, L_{f+1}, f\right) = \exp\left[-\frac{D\left(L_f, L_{f+1}\right)}{m_1 \cdot g(f)} \right] - \exp\left[-\frac{D\left(L_f, L_{f+1}\right)}{m_2} \right] \tag{13}$$

$$g(f) = \begin{cases} 2, & f = 1 \\ 1 + 2/(f-1), & f \geq 2 \end{cases} \tag{14}$$

where $D(L_f, L_{f+1})$ is the Euclidean distance between the current and next potential fixation location, and m_1 and m_2 are free parameters ($m_1 > m_2$). Equation 13 suppresses the probability of both short and long saccades to fit the distribution of the saccade amplitude from human data. Equation 14 further fine tunes the parameter m_1 according the index of the current fixation. The forms of Equations 13 and 14 were selected by hand.

Memory Capacity. For simplicity, we set the memory of the model to be all-or-none, that is, the model either retains complete information from a fixation or completely forgets it. A model with a memory capacity of M can only integrate information from the past $M-1$ fixations plus the current fixation. The posterior probability target being at location i could be calculated as (see supplemental material for the derivation of this formula)

$$P_{i,T_F} = \frac{p(i) \cdot \exp\left[\sum_{f=\min(1,F-M+1)}^{F} \frac{a^2_{i,L_f} \cdot W_{i,L_f,T_f}}{2 + 2\exp\left(-T_f \cdot k_{i,L_f}\right)} \right]}{\sum_{j=1}^{n_0} \left\{ p(j) \cdot \exp\left[\sum_{f=\min(1,F-M+1)}^{F} \frac{a^2_{j,L_f} \cdot W_{j,L_f,T_f}}{2 + 2\exp\left(-T_f \cdot k_{j,L_f}\right)} \right] \right\}} \tag{15}$$

To find the appropriate memory capacity, we first fit the free parameters with unlimited memory in the model and then evaluated the model with various memory

capacities. We calculated the Bhattacharyya coefficients (BC) of the fixation duration distribution, saccade amplitude distribution, and distribution of fixation distance from the image center for each of the initial 20 fixations and saccades using Equation 18 and then averaged across the ordinal positions. This allowed us to evaluate how the model fit the data over time. We then chose the memory capacity that best fit the data and refit the five free parameters in Equations 10 and 13.

Fitting Model Parameters

Table S1 in the supplemental material shows a summary of the parameters used in the ELM, CTELM, and constraint CTLEM searcher. The ELM searcher does not have free parameters that control the eye movement statistics.

For the CTELM searcher, there are three free parameters in Equation 10 that control the distribution of the fixation duration. The remaining parameters were estimated from experimental data and fixed when fitting the free parameters. The values of the free parameters were estimated by genetic algorithm (Goldberg, 2013) that minimized the sum of the Bhattacharyya distance (DB) between the simulated and experimental histograms of the fixation duration (T_{sim}, T_{exp}):

$$Loss = DB\left(T_{sim}, T_{exp}\right) \tag{16}$$

where the $DB(H_1, H_2)$ of the two histograms H_1 and H_2 is defined as

$$DB\left(H_1, H_2\right) = -\ln\left[BC\left(H_1, H_2\right)\right] \tag{17}$$

$$BC\left(H_1, H_2\right) = \sum_{i=1}^{N} \sqrt{H_1(i) \cdot H_2(i)} \tag{18}$$

Here, N represents the number of bins in the two histograms. The histogram should be calculated as probability. The genetic algorithm was implemented in the Global

Optimization toolbox of MATLAB. We set the population size to 150 and max generations to 25, leaving the other parameters unchanged. For each parameter sample in the population, the model was run for 1000 times to obtain the histogram of fixation durations. We used the Multicore package (https://www.mathworks.com/matlabcentral/fileexchange/13775-multicore-parallel-processing-on-multiple-cores) to run the genetic algorithm on multiple computers. A single fitting session took approximately 5 hours to run on a cluster of 14 computers.

For the constraint CTELM searcher, we fit the five free parameters (Equations 10 and 13) by the genetic algorithm. The cost function is the sum of the Bhattacharyya distance (DB) between the simulated and experimental histograms of the fixation duration (T_{sim}, T_{exp}) and saccade amplitude (A_{sim}, A_{exp}):

$$Loss = DB\left(T_{sim}, T_{exp}\right) + DB\left(A_{sim}, A_{exp}\right) \tag{19}$$

We set the population size to 250 and max generations to 25, leaving the other parameters unchanged. For each parameter sample in the population, the model was run 1000 times to obtain the histograms of the fixation durations and saccade amplitudes.

Simulating Visual Search

The three models were evaluated for 10000 trials to obtain eye movement statistics. We randomly rotate the 400 potential target locations around the image center before each trial to obtain a smooth distribution of fixation locations. The target locations were randomly selected with equal probability in each trial.

To obtain the relationship between target eccentricity and number of fixations needed to find the target, we set the target to appear more frequently around image center during the simulation, so that roughly equal number of trials could be obtained when grouping

simulation trials according to target eccentricity. The three models were evaluated for 10000 trials to obtain eye movement statistics

To obtain the relationship between memory capacity and the model's eye movement metrics, we first assumed unlimited memory capacity in the constraint-CTELM searcher, and fit the five free parameters (equation 10, 13) according to participants' data. We then evaluated the model with memory capacity of 2, 4, ⋯, 14 and infinite fixations for 10000 times each. The TH_{targ} value was set to 0.97. For the simulation result belonging to each capacity, we calculated the histograms of fixation duration, fixation distance and saccade amplitude of the initial 21 fixations and 20 saccades. Then for each of the above three eye movement metrics, we calculated the BC value between histograms of simulation and experiment data corresponding to the same fixation or saccade index. Thus, for each memory capacity value, we obtained 21 BC values of fixation duration, 20 BC values of fixation distance, and 20 BC values of saccade amplitude. Finally, for each eye movement metrics, we averaged the BC values across the fixation or saccade index.

To compare the target detection probability within the first fixation when the target was at screen center between the ELM searcher and CTELM searcher, we simulated the first fixation of the CTELM searcher for 10000 times. The fixation duration was set to be 250 ms. The target was identified if the posterior probability at screen center exceed the target detection threshold at any time within the 250 ms. The target detection threshold was varied between 0.01 to 1, and for each threshold value, we calculated the proportion of trials in which the target was identified.

When we summarized the eye movement metrics, we did not include the last fixations for the fixation duration, nor did we include the initial fixations for the fixation location. We

used BC to quantify the similarity between the histograms of the model's and humans' eye movement metrics. When the histograms are calculated as probabilities and have the same number of bins, BC has a maximum value of 1 if the two histograms are completely the same and a minimum value of 0 if the two histograms have no overlap.

Statistical Analysis

To compare the fixation number between the CTELM searcher (or constraint-CTELM searcher) to the ELM searcher, we first grouped the trials according to target eccentricity by 0.5° step. The average fixation number of the two models belonging to the same group were compared by two tailed two sample t-test with unequal variance assumption (the "ttest2" function in MATLAB with option 'Vartype' set to 'unequal'). Though t-test assumes the samples to be normally distributed, the test is robust with respect to violation in normality if the sample size is large. In our case we could obtain 498 to 692 (mean = 610) simulation trials for each group of eccentricity of each model, so the test should still be effective even if the data is not normally distributed.

Code Accessibility

The code for analyzing eye-tracking data and the visual search models is available upon request to the corresponding author.

Results

Selecting Target RMS Contrast

The participants' correct response rate can be well fitted by a Weibull psychometric function given the target RMS contrast (Figure 1B). Visibility was positively related to the target RMS contrast and can be described by Equation 2 (Figure 1C). The selected target RMS contrast ranged from 0.111 to 0.125, with a mean value of 0.121.

Temporal Course of Target Visibility

The target visibility increased as a function of time after stimulus onset (Figure 1D and Figure 1E). The steady-state foveal visibility was approximately 3.14, which was close to the predefined value of 3.0. The rising speed of visibility varies across the visual field. In the foveal region, the visibility peaked rapidly within the initial 100 ms; however, the peripheral regions rose slower and had lower peak visibility levels (Figure 1E). Notably, at just 2.25° (30% of the image radius) away from the fovea, the target visibility peaked after approximately 400 ms (Figure 1E). Therefore, the participants identified targets faster in the fovea and much slower in the peripheral regions.

Because the visibility map was measured with the target location cue in the detection task, one may question the validity of using it in the visual search task when there is no target cue. We measured the visibility map without a target location cue in another group of 4 participants (Figure S5) and found that the optimal search performance obtained by the ELM searcher was much slower than the participants' performance (Figure S6). Therefore, in theory, a visibility map measured without a target location cue does not allow human-level search performance.

Participants' Spatial Control of Fixation was Inconsistent with the ELM Rule

We first compared the distribution of the fixation location, saccade amplitude, and saccade direction against the ELM rule. The participants' fixation density was doughnut shaped, with fewer fixations in the central part of the image and more fixations at the outer part (Figure 3A). However, there was a fixation "hotspot" located slightly above the image center. The "hotspot" was actually caused by the first saccade from a single participant. Because the stimulus images were statistically symmetric around the center, we choose to

focus more on the overall feature of the distribution instead of the local variation. This doughnut-shaped distribution of fixation locations could be generated by both the ELM and CTELM searcher (Figure 3A); however, an important difference emerged when we split the distribution by fixation order. While the distribution of fixation distances from the image center was relatively stable as the search progressed for the human searcher, the two models progressively fixated at the outer part of the image (Figure 3C). The fit to the participants' fixation distance decreased as the search progressed (Figure 3B).

For the saccade amplitude, the participants' distribution was different from the two optimal searchers' distribution (BC = 0.898 for ELM and BC = 0.900 for CTELM), because the two optimal searchers generated more long saccades than the participants did (Figure 4A). The difference was present as early as the third saccade (Figure 4C), and the disparity between the models' and humans' distributions increased as the search progressed (Figure 4B).

For the saccade direction, the participants preferred to make horizontal and vertical saccades over saccades of oblique directions, whereas the ELM and CTELM searcher showed a weak preference of the vertical saccades (Figure 5A). The difference is larger in the first saccade and after 15 saccades (Figure 5B and Figure 5C). The participants' first saccade was very biased to the upward direction (-90°) and generated the fixation hotspot located above the image center. This was the behavior of a single participant; for simplicity of the model, we chose to ignore this variation. Taken together, our data reveal that participants used a different search strategy than the ELM rule.

CTELM Searcher Constrained by Memory and Saccadic Bias Explained Participants' Spatial Control of Fixation

Because the participants actually searched quite fast (2 fixations slower than the ELM searcher on median, as will be discussed below) and because the overall fixation density could already be captured by the ELM rule, it is possible that the participants' strategy did not deviate excessively from this rule. Considering that the human brain may be subject to various limitations, we tested whether a CTELM searcher constrained by a few limitations (constraint-CTLEM searcher) better explained the participants' data.

To make the model's saccade amplitude distribution better agree with the participants' data, we applied a penalty function when calculating the expected information gain of the next fixation (Equations 13 and 14). However, this limitation alone cannot prevent the model from gradually fixating toward the outer edge of the image. This is because the ELM and CTELM searchers tended to avoid examined locations that became larger around the image center as the search progressed. This implied that the perfect memory of all previous fixations in the ELM and CTELM searcher may be the cause of this problem, and the human brain does not have an unlimited memory capacity.

Therefore, we vary the constraint CTELM searcher's memory capacity and examined how eye movement metrics fit to participants' data as the search progressed. Figure 6 shows that the distribution of fixation distances was indeed sensitive to memory capacity, and the matching to the experimental data peaked when the memory capacity was 8 fixations. The fixation duration and saccade amplitude were less sensitive; however, the fit also peaked at approximately 8 fixations (Figure 6). Therefore, we choose a memory capacity of 8 fixations to be the optimal capacity. When the five free parameters in the constraint CTELM searcher were re-fit with this capacity, the relationship displayed in Figure 6 did not change (Figure S3).

With a memory capacity of 8 fixations, the constraint-CTELM searcher's fixation density was still doughnut shaped but had more fixations around the center of the image (Figure 3A). The model's fixation distance distribution was relatively stable after the third fixation and consistently fit the participants' data (Figure 3B and Figure 3C). The saccade amplitude distribution of the constraint-CTELM searcher was similar to the participants (BC = 0.967). There was still a noticeable discrepancy, as the constraint-CTELM searcher made more short saccades than human participants, but importantly, the fit of the saccade amplitude to the participants' data remained stable as the search progressed (Figure 4B). For the saccade direction, the constraint-CTELM searcher performed similar to the ELM and CTELM searcher (Figure 5A) because we did not introduce the direction bias. However, the fit of the saccade direction as search progressed was slightly better than the CTELM and ELM searcher after the 15th saccade (Figure 5B).

Taken together, the participants' saccadic bias and limited memory were at least partial causes of the suboptimal searching behavior. The eye movement model agreed with the participants' data better when these limitations were considered.

CTLEM and Constraint-CTELM Searcher Explained Participants' Overall Fixation Duration

Both the CTELM and constraint-CTELM searcher generated similar overall distributions of fixation durations compared to the participants (BC = 0.996 for both models, Figure 7A); however, both models could not fit the first fixation duration's distribution very well (Figure 7B and Figure 7C). The fitting of the fixation duration as the search progressed was worse for the CTELM searcher than the constraint-CTELM searcher, especially after the 14th fixation (Figure 7B). This was not because the CTELM searcher was unable to produce

sufficient trials with more than 14 fixations, as we obtained more data for up to 18 fixations from the CTELM searcher compared to the participants for calculating the histogram. Taken together, both models generated similar overall fixation duration distributions; however, they could only partially explain the participants' changes in fixation duration as the search progressed.

Constraint-CTELM Searcher Better Explained Participants' Search Performance

With the appropriate choice of target detection threshold TH_{targ}, the three models' correct response rate (ELM: 88.0%, CTELM: 87.6%, and constraint-CTELM: 87.5%) was similar to the participants' correct response rate (87.9%). All three models needed a lower median number of fixations (7 fixations) to find the target compared to the participants (9 fixations). The participants' distribution of fixation numbers had a larger tail (maximum of 164 fixations) than the ELM (maximum of 40 fixations) and CTELM (maximum of 35 fixations) searchers; however, this distribution can be captured by the constraint-CTELM searcher (maximum of 176 fixations) (Figure 8A).

The constraint-CTELM searchers show a similar degradation of performance to participants when the target was located more than 5° from the image center, whereas the ELM and CTELM searcher performed much better due to the unlimited memory (Figure 8B and Figure 8C). However, the constraint-CTELM searcher was unable to capture the participants' search inefficiency when the target was less than 3° away from the image center (Figure 8B and Figure 8C); thus, there may be other factors limiting the participants' performance. Taken together, the constraint-CTELM searcher explained the participants' search inefficiency better than the CTELM and ELM searcher.

CTELM and constraint-CTELM Searcher Detect Target More Reliably Than the ELM

Searcher

An unexpected discovery worth noting was that the CTELM and constraint-CTELM searcher, on average, performed significantly better than the ELM searcher when the target was less than 3° from the image center (Figure 8C, Table S2). Importantly, if the target appeared at the image center, the CTELM and constraint-CTELM searcher could detect the target within the first fixation for 97% of the trials, while the ELM searcher achieved this for only 53% of the trials. The reason was not that the former two models on average fixated longer than 250 ms (which the ELM searcher assumed) in the first few fixations (Figure 7C), because a modified-ELM searcher considering the change of average fixation duration still performed worse when the target was close to the image center (Figure S7, Table S3). The actual reason was that the ELM searcher only considered posterior probability at exactly 250 ms after the fixation onset, while the CTELM and constraint-CTELM considered the whole diffusion process. The posterior probability could first rise above the threshold but drop below it at 250 ms, and in this case the ELM searcher would fail to detect the target, but the CTELM searcher would not. Figure 9 shows that for any target detection threshold, the CTELM searcher always detects the target located at the screen center within the first fixation better than the ELM searcher. Therefore, the better performance of CTELM searcher actually showed the importance of using a continuous-time eye movement model rather than a model with fixed fixation duration.

Discussion

In this research, we first proposed the CTELM searcher, which optimally controls eye movements both in space and time. We showed that the participants' search strategies were suboptimal. However, by biasing the model's saccade amplitude and limiting the model's

memory capacity to 8 fixations, we obtain a model that closely matched human eye movement statistics in visual search. Compared to previous studies, our model no longer uses artificial evidence accumulation process, agrees with human behavior better, and is more biologically realistic as it considers limitations in memory. The importance of modelling the full accumulation process rather than assuming a constant fixation duration was shown in the better performance of the constraint-CTELM searcher than the ELM searcher when the target was closed to image center.

The Control of Fixation Duration

Recently, the "mixed control" hypothesis of fixation duration has received experimental support (Utrecht, 2006; Henderson and Pierce, 2008; Walshe and Nuthmann, 2014) and was adopted in eye movement models of scene viewing (Nuthmann et al., 2010; Trukenbrod and Engbert, 2014). This hypothesis states that fixation duration is controlled by an autonomous timer and modulated by processing demand from the current fixation. The constraint-CTELM searcher used a similar control strategy. The autonomous timer and modulation from the current fixation were implemented as the collapsing bound and the drift-diffusion process, respectively. Fixation will be prolonged if the model cannot determine whether the current fixation location has a target because the probability value at the fixation location will not decrease sufficiently quickly.

The constraint-CTELM searcher's inability to reproduce the participants' first fixation duration suggests that the participants initialized the search differently than the model. Participants may already have a saccade plan before the trial starts; thus, they could make the first saccade earlier, whereas the model had to collect sufficient information before the first saccade. This may also explain why the participants' average search performance

was inferior when the target was near the image center, because efficient saccade planning requires sufficient time to plan (Ghahghaei and Verghese, 2015).

Suboptimal Human Eye Movements

The participants' spatial control of eye movements was suboptimal and showed prominent bias. Considering the importance of vision for survival, finding the possible cause and evolutionary advantage of this suboptimal behavior is crucial to understanding human vision strategies in the real world.

First, the participants preferred short saccades, whereas the ELM and CTELM searcher made more long saccades. The preference is also shown in other visual tasks (Tatler and Vincent, 2009). Interestingly, the preference for short saccades does not affect the median search step performance, as shown by the constraint-CTELM searcher. Favoring short saccades may have adaptive advantages not captured in the current visual search task and models. For example, for random, nondestructive searching, Levy flights (which favors short flights) can optimized the number of targets being found per unit distance traveled (Rosa et al., 1999). Besides, human saccades are not accurate, and making short saccades may also help to avoid subsequent corrective saccades because the saccade landing accuracy decreased for longer saccades (Kerr and Reddix, 1988; Abrams et al., 1989). Another possible advantage is that the short saccade occupies a shorter flight time; thus, the disruption of the visual image during the saccade can be minimized (Bridgeman et al., 1975; Collewijn et al., 1988).

Second, as reported in other studies (Foulsham et al., 2008; Van Renswoude et al., 2016), participants preferred horizontal and vertical saccades over saccades of oblique directions. Because the search images were statistically radially symmetric, the preference

could not be generated by image features. One possible cause of the horizontal preference may be the higher cone density in horizontal regions of the retina (Curcio et al., 1990). Indeed, both our data and the visibility map measured by Najemnik & Geisler (2009) showed horizontal enhancement, though the enhancement was weaker in our data. Najemnik & Geisler (2009) showed that the ELM rule could generate the horizontal saccade bias when the horizontal enhancement of the visibility map was prominent; however, our results from the three models showed that the horizontal saccade bias could not be generated when the horizontal enhancement of the visibility map was weak. Another possible cause of both the horizontal and vertical bias is that the human eye movements have been adapted to the statistics of the natural environment. Due to the Earth's gravity, edges in natural scenes are more likely to lie in the horizontal and vertical directions, and there is often more power in the horizontal and vertical directions than in the oblique directions (Torralba and Oliva, 2003). For the human brain, which has limited energy and computational resources, fixating along the horizontal and vertical edges may have a higher chance of catching important visual information.

Many studies have shown that human eye movements are suboptimal when optimally performing a task requiring effortful computation and careful planning (Araujo et al., 2001; Morvan and Maloney, 2012; Nowakowska et al., 2017). Considering the saccadic bias, we think that the control of the saccade amplitude and saccade direction may use a similar "mixed control" strategy as for fixation durations. Evolution and early life experiences may have provided good heuristics when selecting the saccade direction and amplitude; however, specific demands from the current visual task can further modify these metrics. The advantage of this strategy is that the computational cost of saccade target selection may be

further reduced to selecting a few random numbers, as even the relatively efficient ELM rule requires convolving two maps across the visual field in every fixation. The heuristics may be good enough but not strictly optimal for a particular task. This may explain why participants searched quite fast but were slower than the optimal model in our visual search task.

Memory Capacity of Visual Search

Another novel feature in the constraint-CTELM searcher is the limitation of the memory capacity when integrating information across multiple fixations. This is another cause of suboptimal behavior and may explain why participants sometimes made more than 100 fixations before finding the target. However, it is interesting that adding a memory limitation did not change the median number of fixations to find the target compared to the model with unlimited memory; thus, this may allow humans to reserve memory resources and energy while maintaining high search performance.

The type of short-term memory that we modeled is most similar to the fragile visual short-term memory (FVSTM). FVSTM is thought to be an intermediate stage between iconic visual memory and visual working memory. It has a high capacity of approximately 5-15 items, lasts for at least 4 seconds (Sligte et al., 2008) and has been shown to be tied to the location of the original visual stimulus (Report, 2013), making it particularly suitable for integrating information across multiple fixations to generate a posterior probability map. However, the exact memory capacity in visual search tasks remains controversial. There are reports indicating that visual search has a memory of 3 – 4 items (McCarley et al., 2003), 7−9 items (Beck et al., 2006; Kaunitz et al., 2016), and more than 10 items (Dickinson and Zelinsky, 2007). The predicted memory capacity of 8 fixations of the constraint-CTELM searcher lies within the range of these reported results and thus may serve as additional

evidence in this debate. However, the exact memory capacity requires future investigations and may serve as a test of our model.

Limitations of current study

Although the constraint-CTELM searcher explained a range of behavioral data of visual search, it is still a simple model and does not fully reflect the complexities of the visual system. First, accurately measure the target visibility is challenging, because the 2IFC task may not reproduce the actual attention deployment and pre-saccadic enhancement of visual processing (Rolfs and Carrasco, 2012) happened in visual search, and introduced perceptual learning at measured locations. Second, humans may plan multiple fixations ahead during a fixation (Caspi et al., 2004; De Vries et al., 2014; Hoppe and Rothkopf, 2019), whereas our model only plans one fixation ahead. Third, the implementation of memory is still relatively simple. The model either kept information from a fixation or completely forgot it, whereas human brain memory may gradually decay over time (Hesse and Franz, 2010; Ricker et al., 2016). In addition, multiple types of memory may be involved in visual search (Peterson et al., 2007; Van der Stigchel and Hollingworth, 2018).

Participants also had some eye movement features that the constraint-CTELM searcher could not reproduce. First, the model made more short saccades than did participants. Many short saccades were made when the target's posterior probability at the current fixation location was high but not higher than the target detection threshold. In this case, the model tended to saccade to a nearby location and then saccade back to confirm whether the target was indeed there. Second, as discussed above, participants terminate the first fixation earlier than the model, and thus, they may initialize a trial differently. Nonetheless, we believe that our model can be a starting point for a more complete eye

movement model in the future.

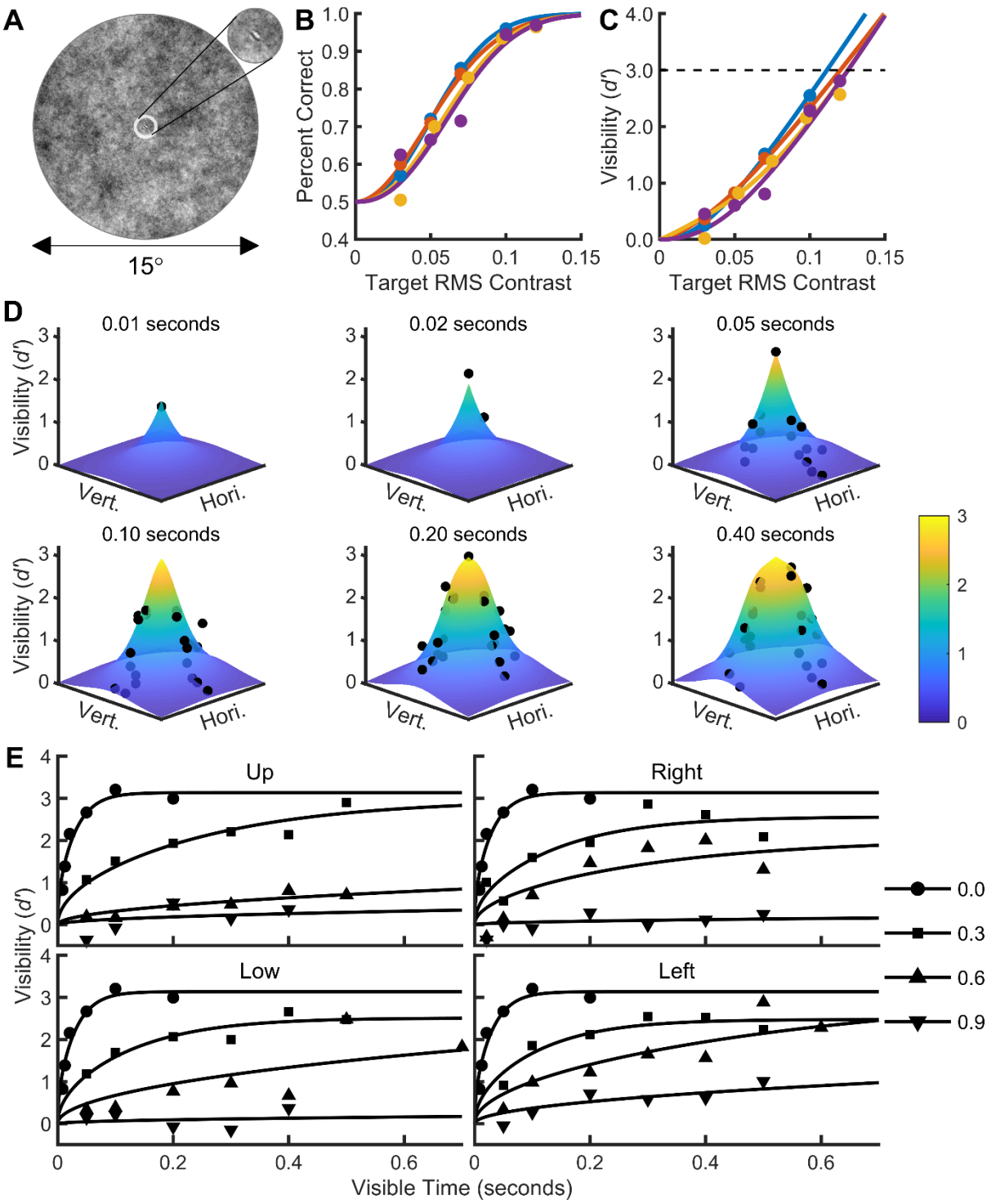

Figure 1. Stimulus image and results of the detection task. **A:** The stimulus image was a 1/*f* noise image (15° diameter, root-mean-squared (RMS) contrast 0.2) with an embedded target. The target was a 6 cycle/degree Gabor grating with a diameter of 0.3 degrees, oriented 45°

counterclockwise from the vertical and windowed by a symmetrical raised cosine (half-height width of one cycle). The area outside the image was set to the mean gray value of the image (0.5, not shown here). **B and C:** Detection rate (B) and foveal target visibility (C) as a function of the target RMS contrast of the four participants. Dots are raw data; different colors (or gray values) represent different participants. In B, the curves are the Weibull function fit to the dots. In C, the curves are calculated from Weibull functions fit to hit rates and the correct rejection rate of the raw data (Equation 1). The intersection between the dashed line ($d' = 3.0$) and the curves is the target RMS contrast that we chose for each participant. This target contrast was used for all subsequent experiments. **D:** Target visibility map after different stimulus exposure times in the detection task. The dots are raw data merged from all participants. The surface is the visibility map function (Equations 3-5) fit to the raw data. Not all locations were tested in these exposure times; thus, the number of black dots may differ in each subplot. Hori/Vert: horizontal/vertical dimension of the search field. Fixation location locates at the center of horizontal and vertical axis. **E:** Temporal course of target visibility at four cardinal directions (shown in four subplots) relative to the fixation location. Each marker represents raw data measured at different relative eccentricities (shown in the legend) merged from all participants. The curves are Equation 3 fitted to the raw data measured at each location. The eccentricity values in the legend are normalized with respect to the radius of the search field. The data measured at the fixation location (eccentricity = 0.0) are shown in every subplot. For clarity, not all measured locations are shown here. See Figure S1 for the complete data.

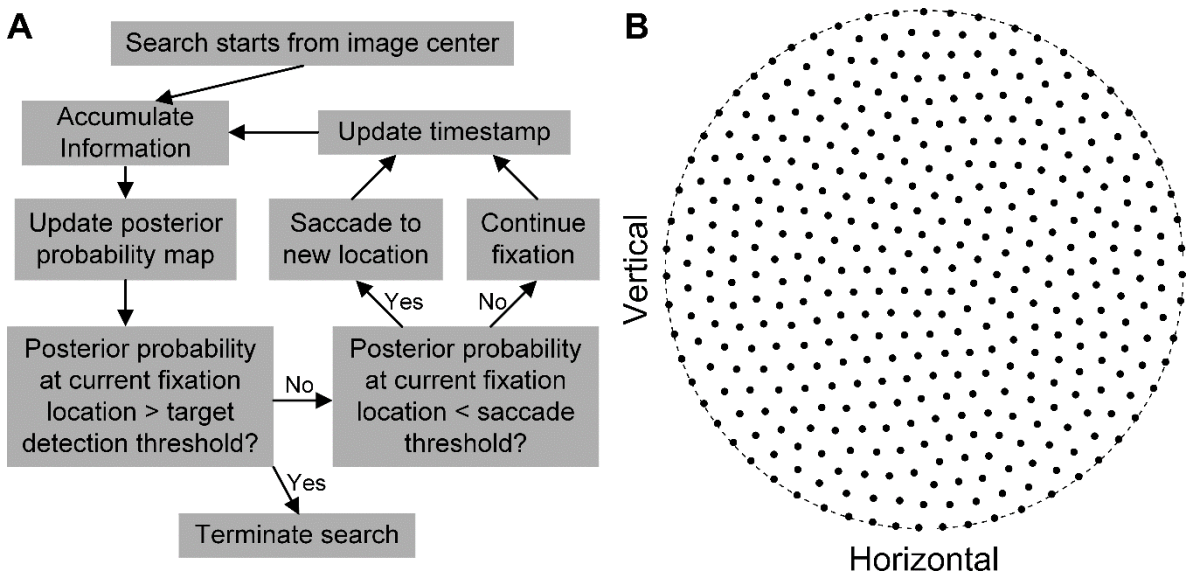

Figure 2. **A:** Schematic diagram of the continuous-time entropy limit minimization (CTELM) and constraint-CTELM searchers. **B:** Locations of the 400 positions (black dots) where the model can fixate. The dashed circle indicates the boundary of the search area. We generated these positions by simulating 400 dots repelling each other. The repelling force was proportional to the inverse sixth power of the distance between two dots. One dot was fixed at the image center and served as the starting position of the visual search.

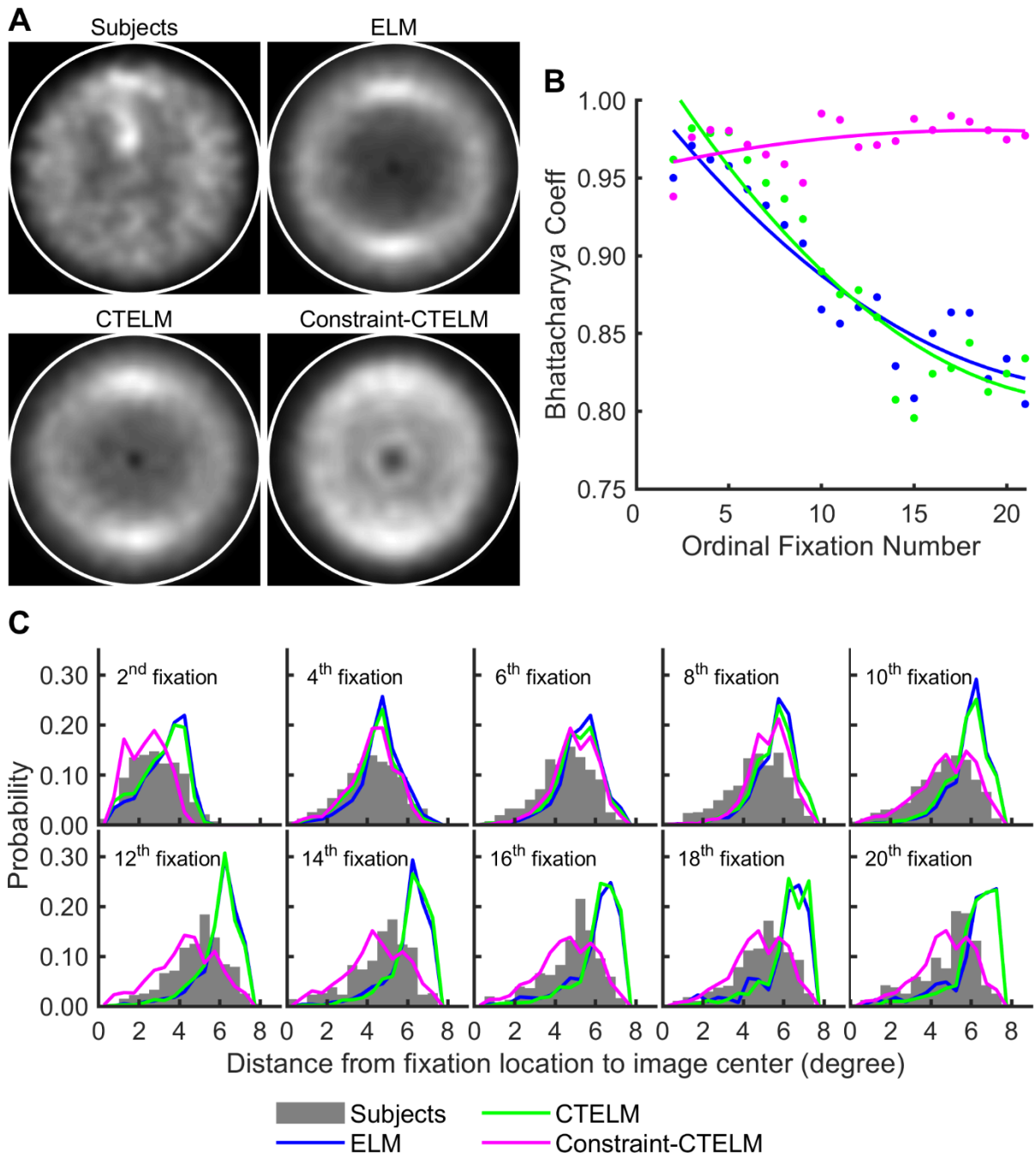

Figure 3. Fixation location distribution of participants and three searchers. **A:** The gray scale represents the relative proportion of fixations located at the displayed location, where lighter means higher density. The relative proportions were obtained by smoothing the scatterplot of the fixation locations by a Gaussian window with a standard deviation of 0.35 degrees (15 pixels) and then scaling so that the maximum value in each subplot equals one. **B:**

Bhattacharyya coefficient between models' and participants' distributions of fixation distances from the image center of the initial 20 fixations after the first saccade. Dots represent raw data, curves represent quadratic functions fitted to the dots. Each color represents the model shown in legend at the bottom of the figure. **C:** Fixation distance distribution selected within the initial 20 fixations after the first saccade. Bars represent participants' data, and lines of different color represent the model shown in legend at the bottom. ELM: the entropy limit minimization searcher. CTELM: the continuous-time entropy limit minimization searcher.

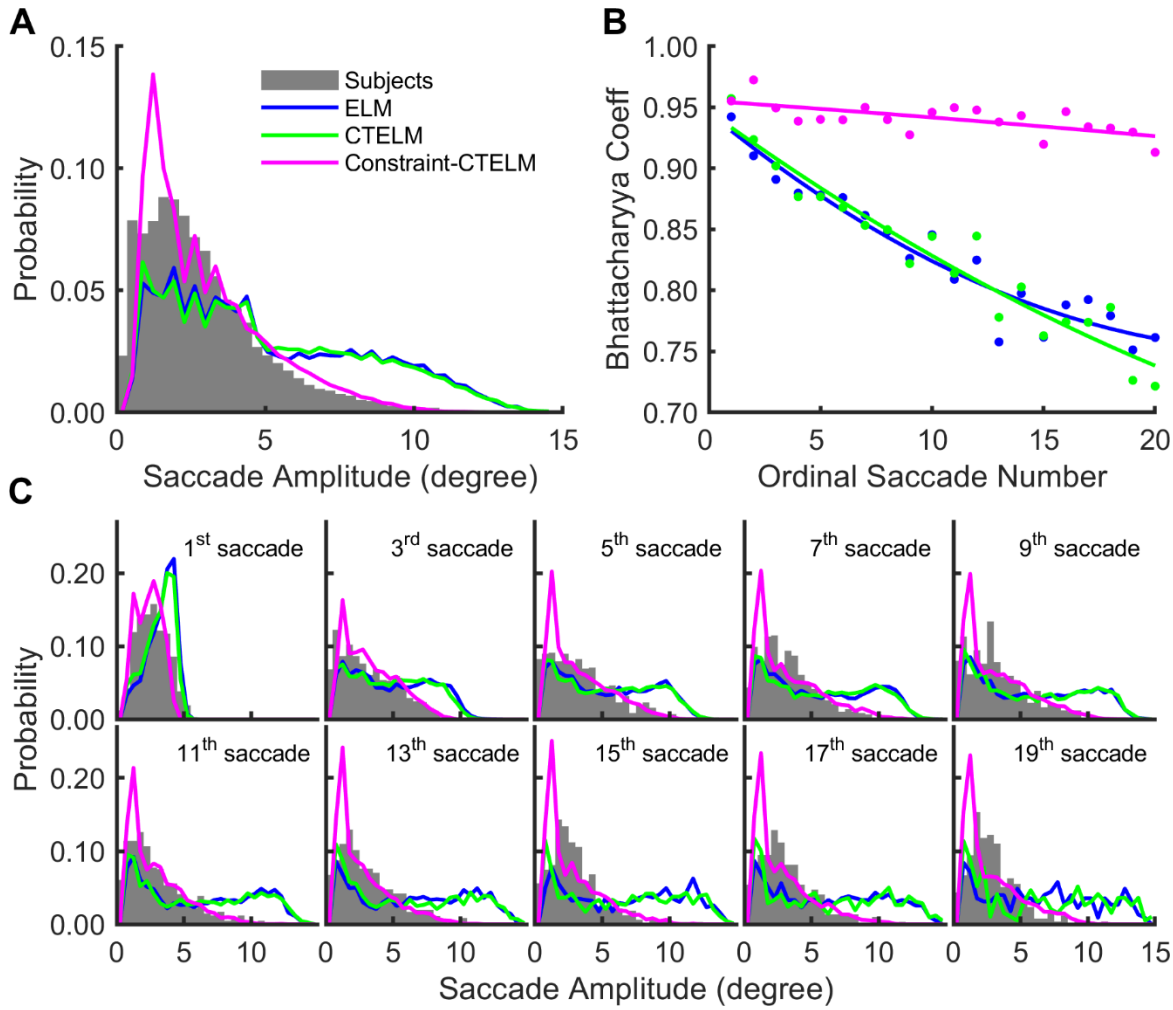

Figure 4. Saccade amplitude distribution of the participants and the three searchers. **A:** Saccade amplitude histogram of all saccades. **B:** Bhattacharyya coefficient between the models' and participants' saccade amplitude distribution of the initial 20 saccades. Dots represent raw data, curves represent quadratic functions fitted to the dots. Each color represents the model shown in legend of panel A. **C:** Saccade amplitude distribution selected within the initial 20 saccades. Bars represent participants' data, and lines of different color represent the model shown in legend of panel A. ELM: the entropy limit minimization searcher. CTELM: the continuous-time entropy limit minimization searcher.

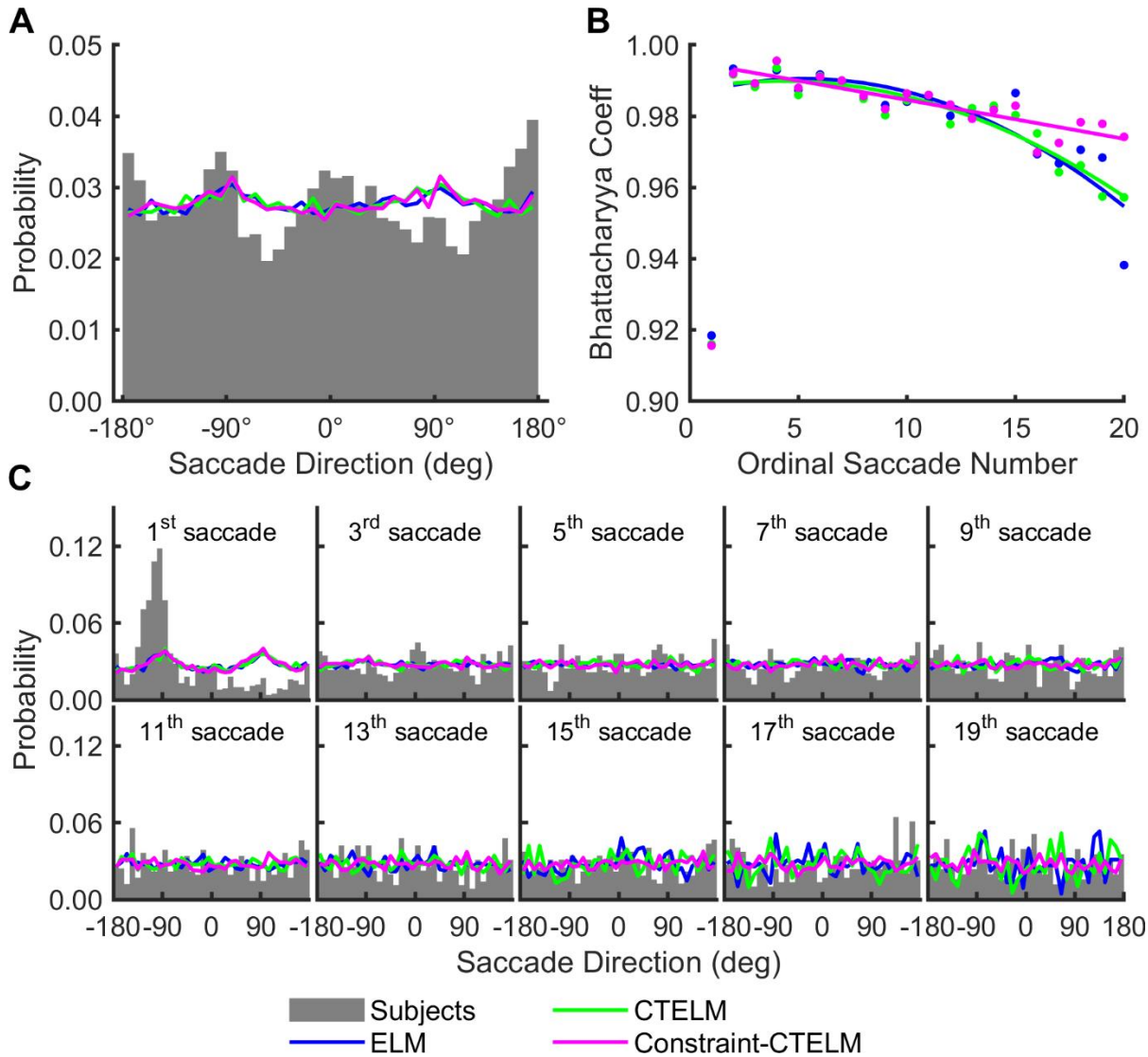

Figure 5. Saccade direction distribution of the participants and the three searchers. The rightward saccade is 0°, and the downward saccade is 90°. **A:** Saccade direction histogram of all saccades. **B:** Bhattacharyya coefficient between the models' and participants' saccade direction distribution of the initial 20 saccades. Dots represent raw data, curves represent quadratic functions fitted to the dots. Each color represents the model shown in legend of panel A. **C:** Saccade direction distribution selected within the initial 20 saccades. Bars represent participants' data, and lines of different color represent the model shown in legend of panel A. ELM: the entropy limit minimization searcher. CTELM: the continuous-time

entropy limit minimization searcher.

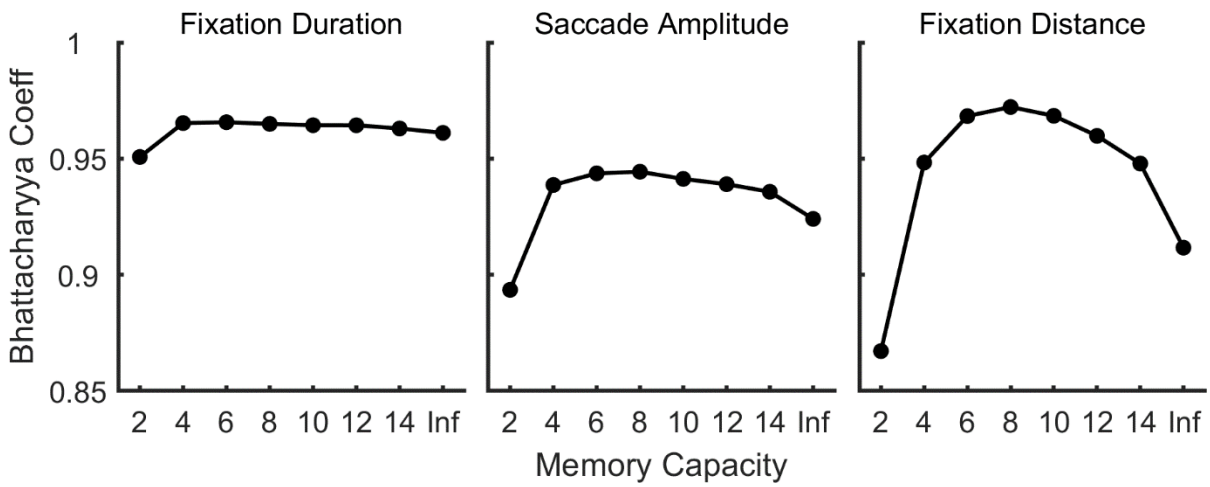

Figure 6. Effect of memory capacity on the average Bhattacharyya coefficients between the sequential fixation duration distribution, saccade amplitude distribution, and the distribution of the fixation distance to the screen center of the simulated data and experimental data. The parameters of the constraint-CTELM searcher used here were fit with an unlimited memory capacity. Each dot is the mean of the Bhattacharyya coefficients separately calculated for each of the initial 20 fixations/saccades. “Inf” in the X axis represents the unlimited memory capacity. CTELM: the continuous-time entropy limit minimization searcher.

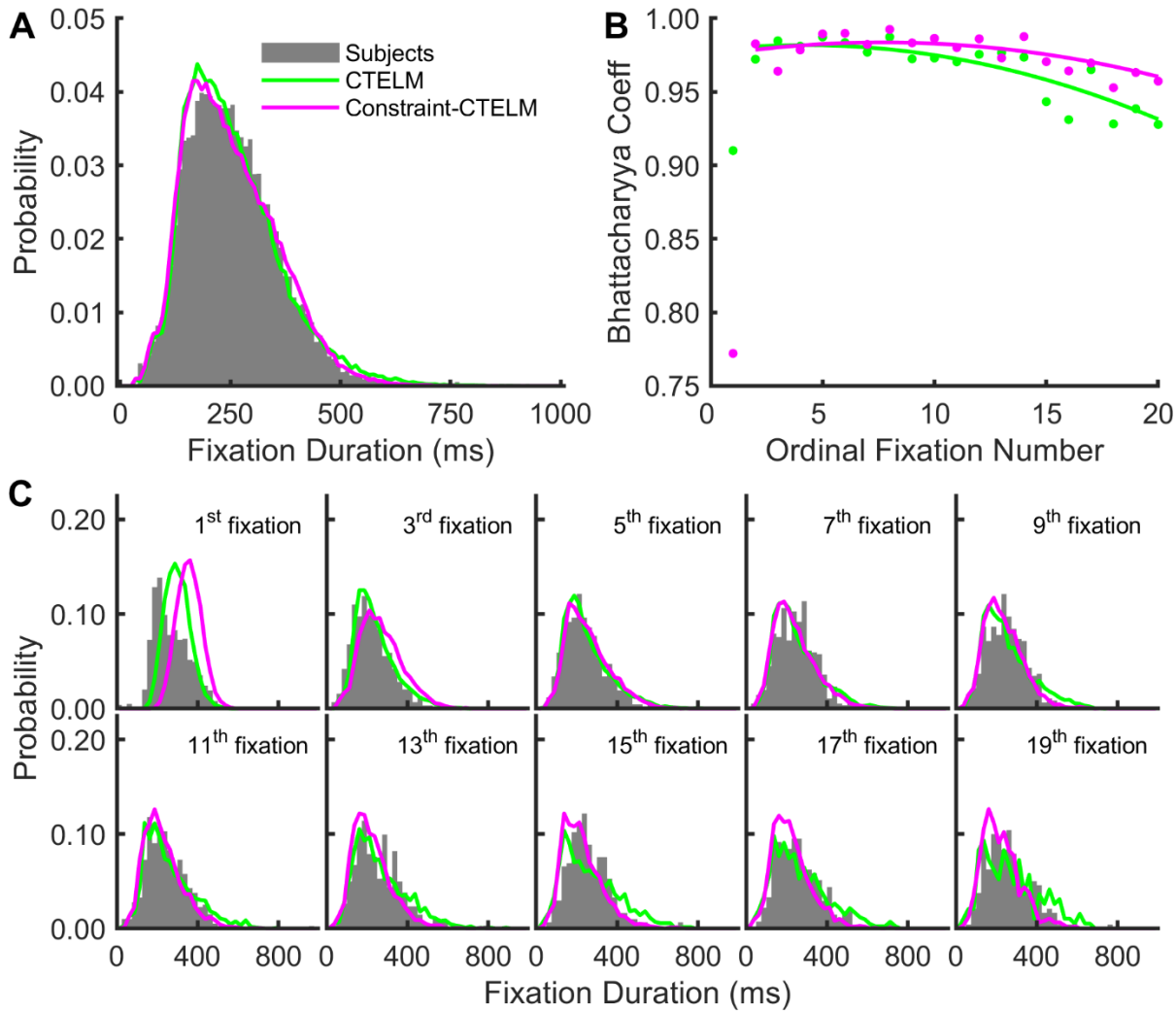

Figure 7. Fixation duration distribution of the participants and the three searchers. **A:** Fixation duration distribution of all fixations, except the last fixation during which the participants and the models made a response. **B:** Bhattacharyya coefficient between the models' and participants' fixation duration distributions of the initial 20 fixations. Dots represent raw data, curves represent quadratic functions fitted to the dots. Each color represents the model shown in legend of panel A. **C:** Fixation duration distribution selected within the initial 20 fixations. Bars represent participants' data, and lines of different color represent the model shown in legend of panel A. The entropy limit minimization (ELM) searcher is not shown here because it did not consider fixation duration. CTELM: the

continuous-time entropy limit minimization searcher.

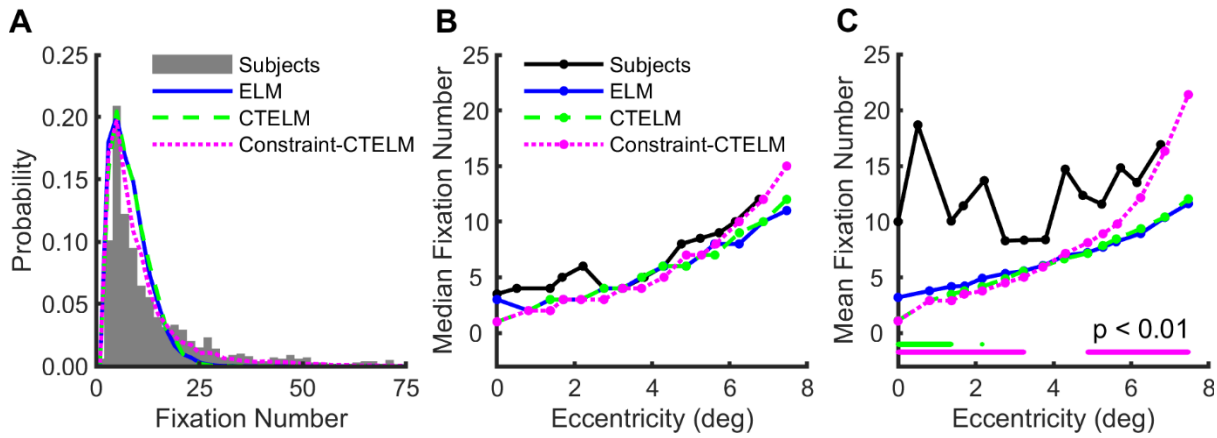

Figure 8. Participants and the three models' search performances. **A:** The distribution of the number of fixations needed to find the target. **B and C:** Median (B) and mean (C) number of fixations to detect the target as a function of the median target eccentricity from the image center. In panel (B) and (C) each data point of the three models represents the median or average fixation number of 498 to 692 (mean = 610) simulation trials. The two bottom lines in panel (C) represent eccentricity range where the average fixation numbers of the CTELM (upper, green) and constraint-CTELM (lower, magenta) searcher were significantly ($p < 0.01$) lower than the ELM searcher. We used the two-sample t-test assuming unequal variance to compare each pair of data points. The exact p-values are shown in Table S2. ELM: the entropy limit minimization searcher. CTELM: the continuous-time entropy limit minimization searcher.

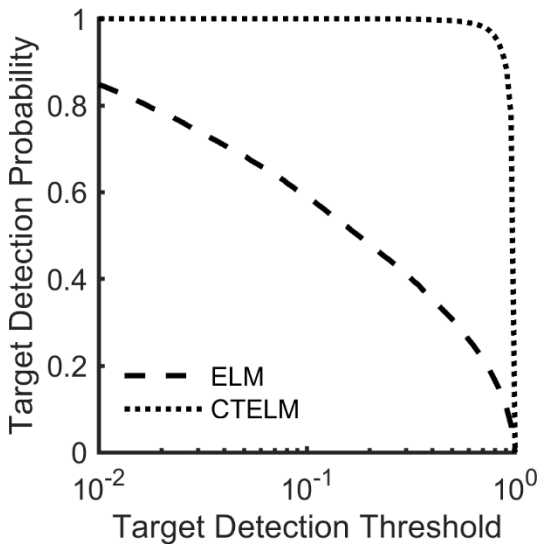

Figure 9. The probability of detecting the target at screen center during the first fixation as a function of target detection threshold of the ELM and CTELM searchers. In the CTELM searcher the fixation duration was also set to be 250 ms, and the target was detected if the posterior probability at fixation location (also the screen center) drifted above the detection threshold at any time within 250 ms. ELM: the entropy limit minimization searcher. CTELM: the continuous-time entropy limit minimization searcher.

References

Abrams RA, Meyer DE, Kornblum S (1989) Speed and Accuracy of Saccadic Eye Movements : Characteristics of Impulse Variability in the Oculomotor System. J Exp Psychol Hum Percept Perform 15:529–543.

Araujo C, Kowler E, Pavel M (2001) Eye movements during visual search : the costs of choosing the optimal path. Vision Res 41:3613–3625.

Beck MR, Peterson MS, Vomela M (2006) Memory for where, but not what, is used during visual search. J Exp Psychol Hum Percept Perform 32:235–250.

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436.

Bridgeman B, Hendry D, Stark L (1975) Failure to detect displacement of the visual world during saccadic eye movements. Vision Res 15:719–722.

Bruce CJ, Goldberg ME, Bushnell MC, Stanton GB (2017) Primate frontal eye fields. II. Physiological and anatomical correlates of electrically evoked eye movements. J Neurophysiol 54:714–734 Available at: https://doi.org/10.1152/jn.1985.54.3.714.

Caspi A, Beutter BR, Eckstein MP (2004) The time course of visual information accrual guiding eye movement decisions. Proc Natl Acad Sci 101:13086–13090.

Clarke ADF, Green P, Chantler MJ, Hunt AR (2016) Human search for a target on a textured background is consistent with a stochastic model. J Vis 16:4.

Collewijn H, Erkelens CJ, Steinman RM (1988) Binocular co-ordination of human vertical saccadic eye movements. J Physiol 404:183–197.

Curcio CA, Sloan KR, Kalina RE, Hendrickson AE (1990) Human photoreceptor topography. J Comp Neurol 292:497–523.

De Vries JP, Hooge ITC, Verstraten FAJ (2014) Saccades Toward the Target Are Planned as

Sequences Rather Than as Single Steps. Psychol Sci 25:215–223.

Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134:9–21 Available at: http://www.ncbi.nlm.nih.gov/pubmed/15102499.

Dickinson CA, Zelinsky GJ (2007) Memory for the search path: Evidence for a high-capacity representation of search history. Vision Res 47:1745–1755.

Dimigen O, Sommer W, Hohlfeld A, Jacobs AM, Kliegl R (2011) Coregistration of eye movements and EEG in natural reading: Analyses and review. J Exp Psychol Gen 140:552–572.

Engbert R, Nuthmann A, Richter EM, Kliegl R (2005) Swift: A dynamical model of saccade generation during reading. Psychol Rev 112:777–813.

Fischer B, Ramsperger E (1986) Human express saccades: effects of randomization and daily practice. Exp Brain Res 64:569–578.

Foulsham T, Kingstone A, Underwood G (2008) Turning the world around: Patterns in saccade direction vary with picture orientation. Vision Res 48:1777–1790.

Ghahghaei S, Verghese P (2015) Efficient saccade planning requires time and clear choices. Vision Res 113:125–136 Available at: http://dx.doi.org/10.1016/j.visres.2015.05.006.

Gluth S, Rieskamp J, Buchel C (2012) Deciding When to Decide: Time-Variant Sequential Sampling Models Explain the Emergence of Value-Based Decisions in the Human Brain. J Neurosci 32:10686–10698.

Goldberg DE (2013) Genetic algorithms in search, optimization, and machine learning, 1st ed. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.

Halverson T, Hornof AJ (2011) A computational model of active vision for visual search in

human-computer interaction. Human-Computer Interact 26:285–314.

Henderson JM, Pierce GL (2008) Eye movements during scene viewing: Evidence for mixed control of fixation durations. Psychon Bull Rev 15:566–573.

Hesse C, Franz VH (2010) Grasping remembered objects: Exponential decay of the visual memory. Vision Res 50:2642–2650 Available at: http://dx.doi.org/10.1016/j.visres.2010.07.026.

Hoppe D, Rothkopf CA (2019) Multi-step planning of eye movements in visual search. Sci Rep 9:1–12 Available at: http://dx.doi.org/10.1038/s41598-018-37536-0.

Horan M, Daddaoua N, Gottlieb JP (2019) Parietal neurons encode instrumental sampling based on decision uncertainty. Nat Neurosci 22:1327–1335 Available at: http://dx.doi.org/10.1038/s41593-019-0440-1.

Kaunitz LN, Rowe EG, Tsuchiya N (2016) Large Capacity of Conscious Access for Incidental Memories in Natural Scenes. Psychol Sci 27:1266–1277.

Kerr PW, Reddix MD (1988) Eye Movement Control During Reading : I . The Location of Initial Eye Fixations on Words. Vision Res 28:1107–1118.

Kleiner M, Brainard D PD (2007) “What’s new in Psychtoolbox-3?” Perception 36 ECVP Abstract Supplement. Perception 36:1–16.

Larson AM, Freeman TE, Ringer R V., Loschky LC (2014) The spatiotemporal dynamics of scene gist recognition. J Exp Psychol Hum Percept Perform 40:471–487.

Macmillan NA, Creelman CD (2005) Comparison (Two-Distribution) Designs for Discrimination. In: Detection Theory: A User’s Guide, pp 165–185.

McCarley JS, Wang RF, Kramer AF, Irwin DE, Peterson MS (2003) How much memory does oculomotor search have? Psychol Sci 14:422–426 Available at:

http://www.ncbi.nlm.nih.gov/pubmed/12930471.

Morvan C, Maloney LT (2012) Human visual search does not maximize the post-saccadic probability of identifying targets. PLoS Comput Biol 8:e1002342.

Munoz DP, Broughton JR, Goldring JE, Armstrong IT (1998) Age-related performance of human subjects on saccadic eye movement tasks. Exp Brain Res 121:391–400.

Najemnik J, Geisler WS (2005) Optimal eye movement strategies in visual search. Nature 434:387–391.

Najemnik J, Geisler WS (2008) Eye movement statistics in humans are consistent with an optimal search strategy. J Vis 8:4.

Najemnik J, Geisler WS (2009) Simple summation rule for optimal fixation selection in visual search. Vision Res 49:1286–1294.

Nowak LG, Bullier J (1997) The Timing of Information Transfer in the Visual System. In: Extrastriate Cortex in Primates (Rockland KS, Kaas JH, Peters A, eds), pp 205–241. Boston, MA: Springer US. Available at: https://doi.org/10.1007/978-1-4757-9625-4_5.

Nowakowska A, Clarke ADF, Hunt AR (2017) Human visual search behaviour is far from ideal. Proc R Soc B Biol Sci 284:1–6.

Nuthmann A, Smith TJ, Engbert R, Henderson JM (2010) CRISP: A Computational Model of Fixation Durations in Scene Viewing. Psychol Rev 117:382–405.

Nyström M, Holmqvist K (2010) An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. Behav Res Methods 42:188–204.

Obara K, O'Hashi K, Tanifuji M (2017) Mechanisms for shaping receptive field in monkey area TE. J Neurophysiol 118:2448–2457.

Pelli DG (1997) The VideoToolbox software for visual psychophysics: Transforming

numbers into movies. Spat Vis 10:437–442.

Peterson MS, Beck MR, Vomela M (2007) Visual search is guided by prospective and retrospective memory. Percept Psychophys 69:123–135.

Prins N, Kingdom FAA (2018) Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the Palamedes toolbox. Front Psychol 9:1250.

Ratcliff R, Frank MJ (2012) Reinforcement-based decision making in corticostriatal circuits: Mutual constraints by neurocomputational and diffusion models. Neural Comput 24:1186–1229.

Report B (2013) Fragile visual short-term memory is an object-based and location-specific store. Psychon Bull Rev 20:732–739.

Ricker TJ, Vergauwe E, Cowan N (2016) Decay theory of immediate memory: From Brown (1958) to today (2014). Q J Exp Psychol 69:1969–1995 Available at: http://dx.doi.org/10.1080/17470218.2014.914546.

Rolfs M, Carrasco M (2012) Rapid Simultaneous Enhancement of Visual Sensitivity and Perceived Contrast during Saccade Preparation. J Neurosci 32:13744–13752.

Rosa S, Conservacion A De, Acg G, Costa ICM De, Especies R, Rica DC, Jose S, Viswanathan GM, Buldyrev S V, Havlin S, Luz MGE, Raposo EP, Stanley HE, Gan R (1999) Optimizing the success of random searches. Nature 401:911–914.

Schiller PH, Sandell JH, Maunsell JH (1987) The effect of frontal eye field and superior colliculus lesions on saccadic latencies in the rhesus monkey. J Neurophysiol 57:1033–1049 Available at: https://doi.org/10.1152/jn.1987.57.4.1033.

Sligte IG, Scholte HS, Lamme VAF (2008) Are there multiple visual short-term memory

stores? PLoS One 3:e1699.

Tatler BW, Brockmole JR, Carpenter RHS (2017) Latest: A model of saccadic decisions in space and time. Psychol Rev 124:267–300.

Tatler BW, Vincent BT (2009) The prominence of behavioural biases in eye guidance. Vis cogn 17:1029–1054.

Thura D, Beauregard-Racine J, Fradet C-W, Cisek P (2012) Decision making by urgency gating: theory and experimental support. J Neurophysiol 108:2912–2930.

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

Trukenbrod HA, Engbert R (2014) ICAT: A computational model for the adaptive control of fixation durations. Psychon Bull Rev 21:907–934.

Utrecht U (2006) Crowding degrades saccadic search performance. Vision Res 46:417–425.

Van der Stigchel S, Hollingworth A (2018) Visuospatial Working Memory as a Fundamental Component of the Eye Movement System. Curr Dir Psychol Sci 27:136–143.

Van Renswoude DR, Johnson SP, Raijmakers MEJ, Visser I (2016) Do infants have the horizontal bias? Infant Behav Dev 44:38–48.

Walshe RC, Nuthmann A (2014) Asymmetrical control of fixation durations in scene viewing. Vision Res 100:38–46 Available at: http://dx.doi.org/10.1016/j.visres.2014.03.012.

Wolfe JM, Gancarz G (1997) Guided search 3.0: A model of visual search catches up with jay enoch 40 years later. In: Basic and Clinical Applications of Vision Science, pp 189–192.

Zelinsky GJ (2008) A Theory of Eye Movements During Target Acquisition. Psychol Rev

115:787–835.

Zelinsky GJ (2012) TAM: Explaining off-object fixations and central fixation tendencies as effects of population averaging during search. Vis cogn 20:515–545.

Zhang S, Lee MD, Vandekerckhove J, Maris G, Wagenmakers EJ (2014) Time-varying boundaries for diffusion models of decision making and response time. Front Psychol 5:1364.