PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

A mathematical model of local and global attention in natural scene viewing

Noa Malem-Shinitski[1]*, Manfred Opper[2], Sebastian Reich[1], Lisa Schwetlick[3], Stefan A. Seelig[3], Ralf Engbert[3]

1 Institute of Mathematics, University of Potsdam, Potsdam, Germany, **2** Department of Artificial Intelligence, Technische Universität Berlin, Berlin, Germany, **3** Department of Psychology, University of Potsdam, Potsdam, Germany

* malem@uni-potsdam.de

Check for updates

OPEN ACCESS

Citation: Malem-Shinitski N, Opper M, Reich S, Schwetlick L, Seelig SA, Engbert R (2020) A mathematical model of local and global attention in natural scene viewing. PLoS Comput Biol 16(12): e1007880. https://doi.org/10.1371/journal.pcbi.1007880

Editor: Jakob H. Macke, Stiftung caesar, GERMANY

Received: April 9, 2020

Accepted: October 23, 2020

Published: December 14, 2020

Copyright: © 2020 Malem-Shinitski et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The raw data can be found in this osf repository - https://osf.io/me2sh/ The processed data and code can be found in this github repository - https://github.com/noashin/local_global_attention_model.

Funding: NMS, MO,SR,LS,SAS,RE have been funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - SFB 1294/1 - 318763901. https://www.sfb1294.de/. The funders had no role in study design, data collection

Abstract

Understanding the decision process underlying gaze control is an important question in cognitive neuroscience with applications in diverse fields ranging from psychology to computer vision. The decision for choosing an upcoming saccade target can be framed as a selection process between two states: Should the observer further inspect the information near the current gaze position (local attention) or continue with exploration of other patches of the given scene (global attention)? Here we propose and investigate a mathematical model motivated by switching between these two attentional states during scene viewing. The model is derived from a minimal set of assumptions that generates realistic eye movement behavior. We implemented a Bayesian approach for model parameter inference based on the model's likelihood function. In order to simplify the inference, we applied data augmentation methods that allowed the use of conjugate priors and the construction of an efficient Gibbs sampler. This approach turned out to be numerically efficient and permitted fitting interindividual differences in saccade statistics. Thus, the main contribution of our modeling approach is two–fold; first, we propose a new model for saccade generation in scene viewing. Second, we demonstrate the use of novel methods from Bayesian inference in the field of scan path modeling.

Author summary

Switching between local and global attention is a general strategy in human information processing. We investigate whether this strategy is a viable approach to model sequences of fixations generated by a human observer in a free viewing task with natural scenes. Variants of the basic model are used to predict the experimental data based on Bayesian inference. Results indicate a high predictive power for both aggregated data and individual differences across observers. The combination of a novel model with state-of-the-art Bayesian methods lends support to our two-state model using local and global internal attention states for controlling eye movements.

and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

The human visual system acquires high-acuity information from a rather small region (the fovea) surrounding the center of gaze [1]. The foveal organization of the visual system has two immediate consequences. First, visual perception of natural scenes depends critically on the control of precise and fast eye movements (saccades) that move regions of interest into the fovea for high-acuity processing. During a typical visual task (e.g., scene viewing or reading), saccades occur at a rate of 3 to 4 per second [2]. Second, the decision process for an upcoming saccade target poses a dilemma: should the observer further exploit the information near the fovea or continue with exploration of other patches within the given scene? The latter problem is critical for scene viewing [3, 4] and relevant to the broader field of cognitive processes in knowledge acquisition [5].

Observers select saccade targets from a priority map [6] that represents objects and regions within a given scene according to their attentional weight. Over the last decades, computational modeling of visual attention for natural scenes [7] resulted in a broad range of successful models [8] of priority maps. These models use feature maps to combine low-level saliency and top-down control. Recently, deep neural network (DNN) models achieved state–of–the–art performances in predicting saliency maps from images [9, 10]. From these advances, the problem of modeling priority maps seems basically solved [11, 12]: for an arbitrary natural image, computational models can generate a prediction of fixation density in experiments with human observers.

The next step in modeling human visual behavior is fundamentally related to the fact that eye movements introduce sequential steps in information processing. Since access to visual information is effectively limited to the fovea, the full sequence of saccadic gaze shifts (scan path) needs to be modeled in order to understand the underlying principles. Understanding how human observers shift their attention while looking at an image requires quantifying the scan paths (Fig 1).

So far, few models for scan path generation and prediction have been proposed. These models can be generally classified into two groups, where one group of models is hypothesis–based and the other is hypothesis–free. The second group includes models which use state of the art deep learning techniques [13, 14]. While these models capture structure present in the data, they provide only very limited insights into the underlying principles of scan path generation. Another critical point for experimental research is that deep learning models require a

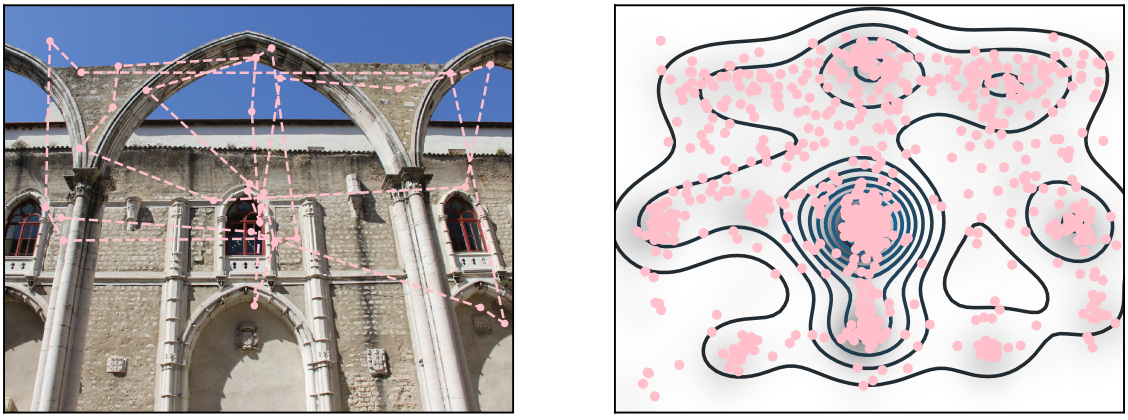

Fig 1. Experimental scanpath and fixations density. *Left*. An image and a scan path. Each dot is a fixation and the dashed line illustrates the saccade. *Right*. The empirical fixation density map as generated by aggregating the fixations from all subjects for a given image.

https://doi.org/10.1371/journal.pcbi.1007880.g001

lot of training data, which are typically unavailable for single observers. Thus, current deep learning approaches do not capture interindividual differences in statistical properties of scan paths.

Hypothesis–based models rely on cognitive and neural assumptions of human perception and oculomotor control that were derived from known biological mechanism and well-established experimental effects [15–19]. Thus, the key goals of hypothesis–based models are (i) to implement these assumptions in a fully quantitative way and build a generative model, (ii) to fit the model to experimental data for hypothesis testing (statistical inference), and, finally, (iii) to provide explanations for interindividual differences in experimental data sets [20].

In the current study, we introduce a new model which belongs to the class of hypothesis–based models. As stated above, we do not model the construction of a priority map. Rather, we address the question of how saccades are generated given a specific static priority map, and use the experimental priority map as input to our model. Our central hypothesis is that the generation of scan paths is based on switching between two internal states of local versus global attention. In this view, the generation of each saccade is a decision process, where the observer has to choose between following the local attention map, and perform a short saccade for staying in the immediate surrounding of the current fixation and the global attention map, and perform long saccade to explore a new region of the visual environment. We assume that the decision is based on the information currently available to the observer. Specifically, this assumption translates into a higher probability of following the local attention map if the ratio of priority values of the current fixated location and the previously fixated location is high. This hypothesis follows an assumption that the area next to a location with high priority also has high priority. This assumption is valid for natural images used in this work.

In implementing a model which changes between local and global attention mode, we continue the line of work that started already in 1976 with the work of Frost and Pöppel [21]. In work by Unema et al. [22] and Helmert et al. [23] it was argued that global attention mode is limited to the beginning of the scan path. Later work by Tatler et al. [15] showed that this is not the case. Thus our model allows the choice between a local or global attention policy throughout the entire viewing period.

Our model might also be interpreted in the context of the Exploration–Exploitation dilemma [24, 25]. In this framework, a decision for choosing an upcoming saccade target is based on two alternatives: Should the observer further exploit the information near the current gaze position or continue with exploration of other patches within the given scene? Hence, a saccade that is generated by the local attention map may correspond to an exploitation step, and a saccade that is generated by the global attention map may correspond to an exploration step. This approach does not take into consideration saccades that return to a previously visited location, and could be interpreted as an exploitation step, which is inline with the our model this phenomenon.

The idea of exploration and exploitation intentions in visual behavior was studied previously by Gameiro et al. [3]. Their work demonstrated experimentally that the tendency for exploration or exploitation, measured by saccade amplitude and fixation duration, depends on size and spatial properties of the stimulus. The characterization of the exploratory or exploitative tendencies was done using the statistics of the entire scan paths.

Different from the approach taken by Gameiro et al. [3], we analyze the individual saccades rather than entire scan paths. Our generative model tags each saccade as either a step that follows the local attention map, or a step that follows the global map.

In the current study, we model natural scene viewing which cannot be directly associated with a reward. Thus our model does not have the notion of value when choosing which policy to follow. As the terminology of Exploration and Exploitation is associated very often with the

notion of the value of the decision, we avoid this terminology, and use the terminology of Local and Global Attention policies, rather than Exploration and Exploitation policies.

We aim at a minimal model to keep computations efficient and to facilitate interpretations of the model behavior, and we do not expect it to captures all the known features of human vision. A critical component of our approach is the application of Bayesian statistics to fit the model to experimental scan path data. We use the fitted model to quantify how well the model describes the experimental data. Further we test different variations of the model, which correspond to different hypotheses, to determine which hypothesis corresponds best to the experimental data.

In the next section we describe the details of our basic model and explain the computation of the likelihood function as a fundamental tool for statistical inference. We construct the model in a modular way and relate each part to one of the assumptions we would like to investigate. Next, we describe the process of fitting the model parameters to experimental data. In the Results, we compare several statistics of simulated data to the statistics of the experimental data. We also analyze different variants of the basic model and quantify how well each one of them describes the data using the model's likelihood function. We close with the Discussion of our results in the context of current problems in understanding scan path generation during scene viewing.

Materials and methods

The local and global attention for scan path generation

Our theoretical investigation of local and global attention in saccadic behavior is based on the implementation of a probabilistic generative model. The static viewer independent priority map for saccadic selection [6] is thought to be the combined result of early visual processing or *saliency* [7] and top-down cognitive control. While various models for the computation of static priority maps exist, we extend the modeling approach to the generation of scan paths for a given static saliency map. For simplicity, we use the time-averaged fixation density [18] as an approximation of the saliency of a given image.

The static saliency map is a function $s(z) : \mathbb{R}^2 \mapsto \mathbb{R}^+$ with $z = (x, y)$ being a location in an image and $s(z)$ being the probability of an average viewer to fixate this location (its saliency). As mentioned above, we approximate the saliency map by the experimentally-observed fixation density and we use $s(z)$ or s_z to refer to the saliency map or the fixation density of the image at location z.

Generally, scan paths are sequences of fixation locations and fixations duration. In this work we model only the spatial properties of gaze control. We account only for the temporal ordering of the fixations and do not model the fixation duration. In these settings, a scan path is written down as $Z = \{z_1, z_2, \ldots, z_t, \ldots, z_T\}$ with T being the number of fixations in the scan path and z_t being the location of the tth fixation.

We begin constructing our model by assuming that the saccade generation process is a second order Markov process, which means that the probability $p(z = z_t)$ of fixating on a specific location z_t at time step t depends only on the location of the fixation at time $t - 1$ and the fixation at time $t - 2$. The probability of a full scan path is written as

$$p(Z) = p(z_1)p(z_2)\prod_{t=3}^{t=T} p(z_t|z_{t-1}, z_{t-2}). \quad (1)$$

The choice of the second order Markov process reflects our hypothesis regarding the scan path generation and will become clear in the upcoming paragraphs. In principle, it is possible

to construct a simpler model which corresponds to first order Markov process. This would correspond to slightly different assumptions regarding the scan path generation and we refer to such a model in the section discussing simplified models.

We describe the probability of the next fixation being z_t given that the previous two fixation location were z_{t-1} and z_{t-2} in terms of competing local and global attention policies:

Local attention. The next fixation location is chosen close to the current fixation location following a Gaussian distribution around the current fixation location with covariance ϵ, normalized over the entire image. This can be written as

$$p_{\text{local}}(z_t|z_{t-1}) = \frac{n(z_t; z_{t-1}, \epsilon)}{\sum_{z'} n(z'; z_{t-1}, \epsilon)} \tag{2}$$

where $n(z_t;z_{t-1}, \epsilon)$ is a Gaussian density with mean z_{t-1} and covariance $\epsilon = \begin{pmatrix} \epsilon_x & 0 \\ 0 & \epsilon_y \end{pmatrix}$.

Global attention. A potential implementation is that the next fixation location is chosen randomly from the static saliency map of the image. This policy leads to very large saccade amplitudes which are known to be less probable [26]. To integrate this prior regarding the saccade amplitudes knowledge into the model—instead of choosing the next fixation location from the the saliency map, we modulate the saliency map by a Gaussian distribution, which gives a higher weight to areas of high saliency which are closer to the current location.

This approach results in the following expression for the global attention strategy

$$p(z_t|z_{t-1}) = \frac{s(z_t)n(z_t; z_{t-1}, \xi)}{\sum_{z'} s(z')n(z'; z_{t-1}, \xi)} \tag{3}$$

with ξ a diagonal covariance matrix similarly to ϵ, $\xi_x > \epsilon_x$ and $\xi_y > \epsilon_y$.

Having Eq (3) as the global attention policy may result in short saccades similar to the ones generated by the local attention policy when the current fixation is in a high priority area. A solution is to create a repulsion mechanism that forces the saccades generated by the this policy to be of at least a certain length. This is achieved by the following expression

$$p_{\text{global}}(z_t|z_{t-1}) = \frac{\max(s(z_t)n(z_t; z_{t-1}, \xi) - n(z_t; z_{t-1}, \epsilon), 0)}{\sum_{z'} \max(s(z')n(z'; z_{t-1}, \xi) - n(z'; z_{t-1}, \epsilon), 0)}. \tag{4}$$

To avoid negative values for the likelihood we take the maximum between the subtraction and 0. Fig 2 visualizes the two distributions formulated in Eqs (2) and (3).

Our assumption is that each fixation is chosen either from the local attention map described in Eq (2) or the global attention map described in Eq (4). This can be represented as a mixture model

$$p(z_t|z_{t-1}, \rho) = \rho\, p_{\text{local}}(z_t|z_{t-1}) + (1 - \rho)p_{\text{global}}(z_t|z_{t-1}). \tag{5}$$

The model parameter ρ describes the tendency to perform a step following either the local or global attention map. It can be fixed based on prior knowledge or inferred from the experimental data. If $\rho > 0.5$ then the probability for a local step is larger than for a global step for every saccade.

Next we include in our model the assumption that ρ changes depending on the fixation location. We use the notation ρ_t to indicate that the fixation z_t was generated based on ρ_t. Importantly, this notation does not imply that ρ_t is necessarily a function of z_t.

We assume that the decision whether to the local or global attention maps depends on the ratio between the priority values of the current and previous fixated locations. The result is that the viewer is more likely to make a local step if the saliency value of the current fixated

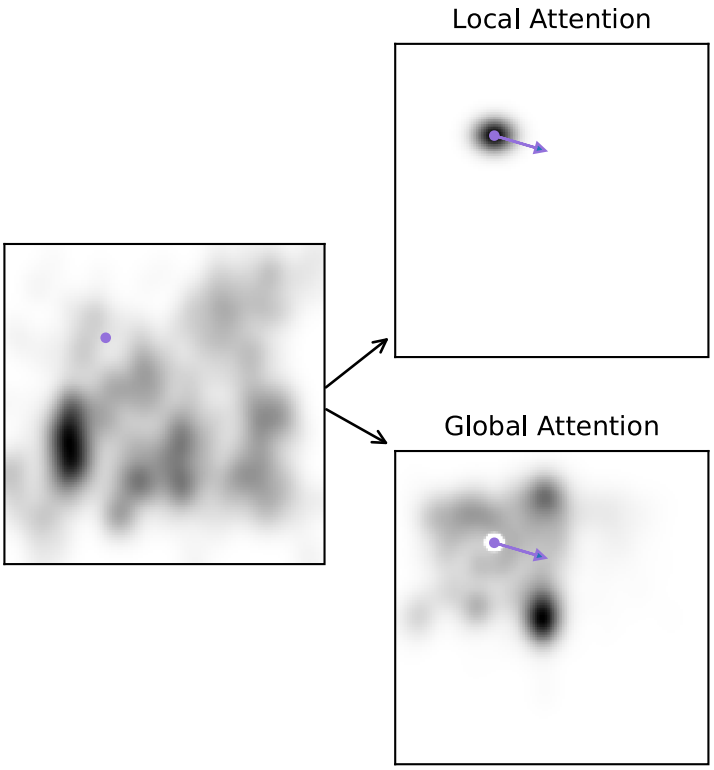

Fig 2. Local and global attention maps. On the left is an example of an empirical saliency map, the dot indicates a fixation location. On the right are the probability maps generated by either the local attention (upper panel) or the global attention policy (lower panel). The arrow indicates a saccade.

https://doi.org/10.1371/journal.pcbi.1007880.g002

location is higher than the saliency value of the previous fixated location. We include this in the model with the following expression for ρ_t

$$\rho_t = \sigma(f(s)) = \frac{1}{1 + \exp(-f(s))} \tag{6}$$

with

$$f(s) = b\left(\frac{s_{t-1}}{s_{t-2}} - s^o\right) \tag{7}$$

with $s_{t-1} = s(z_{t-1})$ and b and s^o being scalar variables.

Combining Eqs (1) and (5), the model likelihood is written as

$$p(Z|\Theta) = p(z_1)p(z_2)\prod_{t=3}^{t=T}(\rho_t\, p_{\text{local}}(z_t|z_{t-1}) + (1-\rho_t)p_{\text{global}}(z_t|z_{t-1})) \tag{8}$$

with model variables $\Theta = \{\epsilon, \xi, b, s^o\}$. Here, we chose to sample the first and second fixation from the empirical static saliency map such that $p(z) = s(z)$.

Fig 3 presents a scan path generated by our model given a particular saliency map, along side a scan path recorded experimentally from a viewer viewing the image corresponding to the saliency map.

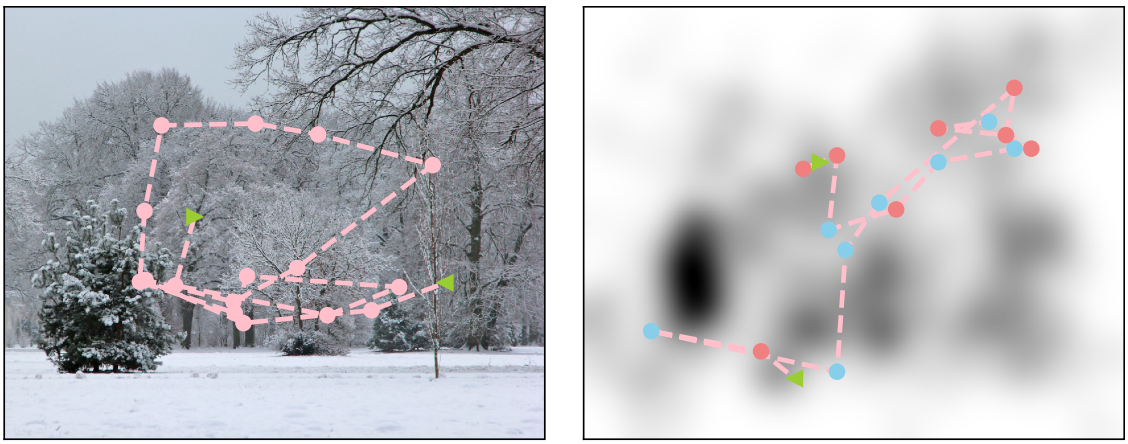

Fig 3. Experimental and simulated scan paths. *Left*. An image and a scan path recorded from a human observer. *Right*. The experimental static saliency map and a scan path generated by the Local and Global Attention Model. The green arrow pointing right represents the second randomly selected fixation location z_2. The green arrow pointing left represents the last fixation in the scan path. The blue dots are fixations that were generated from a global attention step and the pink dots are fixations that were generated from a local attention step. The experimental data shows clearly the phenomenon of saccadic momentum which is not captured by the model. This is further discussed in the Results and Discussion.

https://doi.org/10.1371/journal.pcbi.1007880.g003

Simplified models

To test the different assumptions behind our full model described above, we construct three simpler models and compare their performances to the performance of the full model in the Results. To construct the models we remove one by one the assumptions on which the model is based. This results in the following competing models:

Local choice model. Eq (6) describes the assumption that the decision between two attentions maps depends on the ratio between the priority value of the current fixation location and the priority value of the previous fixation location. A competing assumption would be that the decision depends only on the priority value of the current fixation location. In this case we keep the model the same and only change $f(s)$

$$f(s) = b(s_{t-1} - s^o). \tag{9}$$

Fixed choice model. We test the assumption that the decision between the modes does not depend on the saliency value of previous fixation. In this simplification of the model, rather than having $\rho_t = f(z_{t-1}, z_{t-2})$ we have a fixed probability to chose each policy with $\rho_t = \rho$.

Local saliency model. Last, we challenge the approach of two competing modes. In this variation of the model, each fixation is generated from a modulation of the empirical saliency map with a Gaussian around the current fixation location. This corresponds to the following fixation location likelihood

$$p(z_t|z_{t-1}) = \frac{s(z_t)n(z_t|z_{t-1}, \xi)}{\sum s(z')n(z'|z_{t-1}, \xi)} \tag{10}$$

In the next section we describe the inference process of the full Model. As the three models described above are simplified versions of the full model we do not describe their corresponding inference processes as they can be easily derived from the inference of the full model.

The inference process

Our approach is based on experimental results and we derive the model parameters from observed data in a Bayesian framework. This approach allows us to include prior knowledge regarding the different model parameters based on known spatial features of scan paths. It also allows us to obtain distributions over the model parameters, rather than point estimates, and to compare different variations of the model via the respective test–data likelihoods.

In the previous section we defined the likelihood of the data. Next, we describe the data augmentation methods which allow us to identify conjugate priors and construct an efficient Gibbs sampler [27] using the full conditional distributions over the model parameters.

The idea behind data augmentation [28] is adding latent variables to the model, which can be considered as unobserved data, in a way that simplifies the inference of the parameters of interest. We use the standard approach and augment the Local and Global likelihood by

$$p(Z, \Gamma|\Theta) = p(z_1)p(z_2)\prod_{t=3}^{T} p_{\text{local}}(z_t|z_{t-1})^{\gamma_t}\, p_{\text{global}}(z_t|z_{t-1})^{1-\gamma_t} \tag{11}$$

with

$$\gamma_t \sim \text{Bern}(\rho_t) = \text{Bern}(\sigma(f(s))) \tag{12}$$

and marginalizing over Γ results in Eq (8).

The augmentation defines a modified generative process for the model. At each time step a variable γ_t is drawn from a Bernoulli distribution with bias ρ_t. If the result is 1 then the next saccade is generated following the local attention mode. If the result is 0, the saccade is generated from the global attention mode. This construction reflects our assumption regarding the cognitive process underlying scan path generation, where each saccade follows either local or global attention mode.

For a simple two–component mixture model with normal distribution, the augmentation described above would have been sufficient for the derivation of a Gibbs sampler [29]. As the model we constructed is more complex, we need to handle the sigmoid link–function in Eq (6) and the non-trivial form of the Global Attention distribution in Eq (3).

With the Sigmoid function in Eq (6) there is no straightforward way to define conjugate priors for the parameters b and s^o which are needed for a Gibbs sampler. To achieve conditional probabilities which are easy to sample from, we augment the model with another set of latent variables w_t, which follow a Pólya-Gamma distribution

$$w_t \sim \text{PG}(1, -f(s)). \tag{13}$$

As described in [30] for the case of logistic regression, the usage of this augmentation scheme results in conditional distributions for b and s^o which are Gaussian and can be sampled from easily. The full derivation of the discussed conditional distributions can be found in the supplementary material.

After adding the two sets of latent variable we can define conjugate priors for the parameters:

$$\epsilon_{x/y} \sim \text{IG}(\epsilon_{x/y}; \alpha_{\epsilon_{x/y}}, \beta_{\epsilon_{x/y}}) \tag{14}$$

$$\xi_{x/y} \sim \text{IG}(\xi_{x/y}; \alpha_{\xi_{x/y}}, \beta_{\xi_{x/y}}) \tag{15}$$

$$b \sim \mathcal{N}(b; \mu_b, \sigma_b) \tag{16}$$

$$s^o \sim \mathcal{N}(s^o; \mu_{s^o}, \sigma_{s^o}) \tag{17}$$

where IG is the Inverse Gamma distribution, and $\mathcal{N}$ is the Gaussian distribution.

The prior distributions described above include hyperparameters. These parameters were chosen and not inferred from the data. The hyperparameters related to the prior distributions over $\epsilon_{x/y}$ and $\xi_{x/y}$ were chosen based on known characteristics of human saccades, such as that typical saccade amplitudes range from 0.5 and up to 40 visual degrees [31]. The hyperparameters related to b and s^o were chosen to be on the same scale of the average $\frac{s_{t-1}}{s_{t-2}}$ from the data. Further, all of the hyperparameters were chosen to induce wide prior distributions.

Combining the likelihood in Eq (8) with the priors defined above, the posterior distribution over the model parameters and the latent parameters is given by

$$p(\Theta, \Gamma, W|Z) \propto p(Z|\Theta, \Gamma, W)p(\Gamma|\Theta)p(W|\Theta)p(\Theta) \tag{18}$$

with

$$p(\Theta) = p(\epsilon)p(\xi)p(b)p(s^o).$$

We can sample easily from the conditional distributions of b and s^o. This is not the case for ϵ and ξ because of the form of p_{global}.

Due to the complex form of the global attention expression in Eq (4), which includes both ξ and ϵ, there is no closed form for the conditional distribution of these parameters. Thus, we resort to a technique known as MCMC within Gibbs [32, 33] and in each iteration of the Gibbs sampler we evaluate the conditional distributions of ξ and ϵ using an Hybrid Monte Carlo step [34], also known as Hamiltonian Monte Carlo.

For further technical details regarding the augmentation and the HMC sampler please see the supplementary material.

The implementation of the models and inference can be found under https://github.com/noashin/local_global_attention_model.

Results

In this work we propose a Global and Local Attention Model for scan path generation. In the previous section we derived the model equations from the basic two modes approach. We described the inference process of our model when applied to experimental data. In this Section, we present the results of the inference process. First, we analyzed the reliability of our procedures by fitting the model to artificial data generated from the model with known parameter values. Next, we fit the model to the experimental data and test the statistics of the data generated from the model against the experimental data. Finally, we quantitatively compare different versions of the model.

Model parameters estimation

As presented in the Methods Section, the inference process includes using an MCMC approach to evaluate the posterior function over the model parameters. This approach is exact in the limit of an infinite number of samples but as we can only use a finite number of samples the result is an approximation of the actual posterior. The distribution of the inferred parameters should concentrate around their real values.

When fitting the model to experimental data it is impossible to know the real values of the model parameters as they do not relate directly to any measurable features of the data. Thus, in

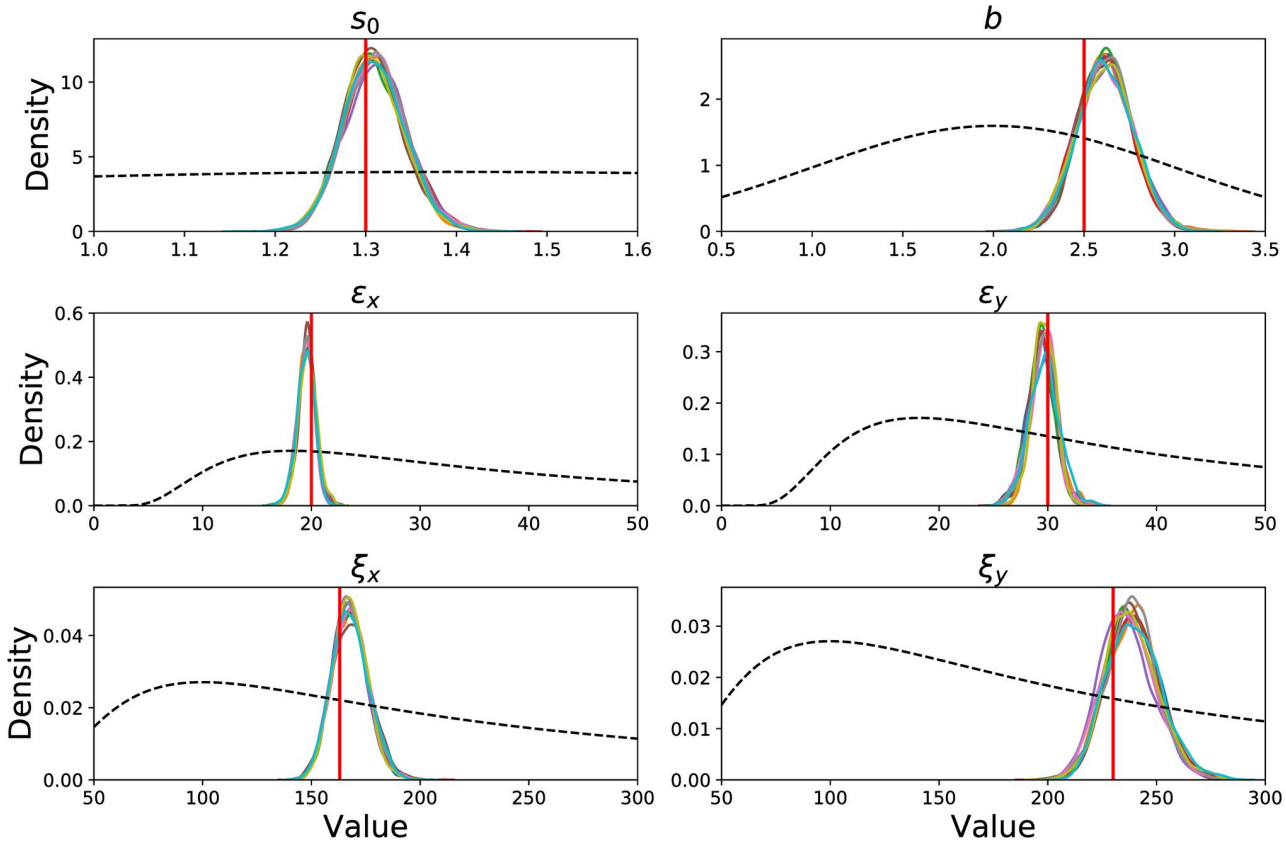

Fig 4. Inference results on simulated data. Model parameter recovery. To test the inference algorithm we fit the model to simulated data with known parameters values. Each panel includes the inferred posterior distribution of each parameter after the inference process. The ten curves present 10 different inference processes starting from different values. The vertical lines are the values with which the data was generated. The black dashed curve is the prior distribution. The plotted densities are not normalized.

https://doi.org/10.1371/journal.pcbi.1007880.g004

order to assess the performance of the inference we use data simulated by the model, in which case we know the exact values used to generate the data. If the inference process is correct we expect the resulting posterior distribution to be concentrated around the ground truth values.

We generated data from our model with the parameter values that were inferred from the experimental data. In order to see whether the inference process will have reasonable results when fitting the experimental data, the size of the generated data set is comparable to the size of the experimental data for one subject.

Fig 4 presents the distribution over model parameters as results from the inference process with data generated by the model. Each of the ten colored curves represents a different inference process started at a different point. As expected all the curves from different runs are similar in shape. The black dashed curves present the prior distribution over the parameters. To test the model we chose the prior distribution so their modes do not overlap with the values used in the data generation. As expected the mode of the inferred parameter distribution is close to the real values used in the data generation which are noted by the vertical solid line.

We tested the model on generated data that have similar properties to the experimental data. Generated scan paths had lengths similar to the lengths of scan paths recorded experimentally. This could have the result that the generated data does not have sufficient information regarding the underlying model parameters and it explains the deviation of the distribution mode from the true parameter values.

Model performance on experimental data

Our model was derived from a set of hypotheses regarding the cognitive process of saccade generation. In order to test the validity of the model, and of the corresponding hypotheses, we fit the model to the data, simulate new data using the model and check whether the features of the simulated data correspond to the features of the experimental data.

The data set used here includes the scan paths of thirty five human observers performing a memorization task over thirty natural images. The same data set was used before to evaluate other scan path models [18, 20]. The participants were presented with an image for 10 seconds and were instructed to explore the scene for a later memory test. The acquisition of the data was carried out in accordance with the Declaration of Helsinki, and informed consent was obtained for experimentation by all participants. Data from three subjects were excluded as the inference process did not converge. The data can be found under https://osf.io/me2sh/.

We fit a separate model for each subject, while using the same prior hyperparameters for all fitted models. We want to test whether the model captures subjects' tendencies that generalize over images. We use the k-fold cross-validation method with $k = 5$. All the reported quantitative results in this section are obtained from the test data averaged over the different folds.

Saliency map recovery. Since the goal of our model is to produce a scan path for a given saliency map, the model needs to recover the empirical saliency map from experimental data. Fig 5 presents the comparison between empirical saliency maps and the fixation locations density of data generated by the model. We used three different empirical saliency maps from the test-set for simulation of the full Local and Global Attention model and generated data from all the models fitted to the different subjects. The contour plot includes the density of the aggregated data, and the density is the empirical saliency map. Qualitatively, as expected, the fixation density of the data generated by the model, matches the empirical saliency maps.

Saccade amplitude. The Local and Global Attention Model was designed to capture the different saccade amplitudes generated by subjects while observing an image in a free viewing task. To estimate the model's performance we compare the amplitudes of the empirical saccades with the amplitudes of the saccades generated by the model. The comparison is done both at a population level and for each subject separately.

Fig 6 compares the empirical saccade amplitude density with the saccade amplitude density of the scan paths that were generated by the full Local and Global Attention Model and the simplified versions presented previously. The density presented is over the entire population

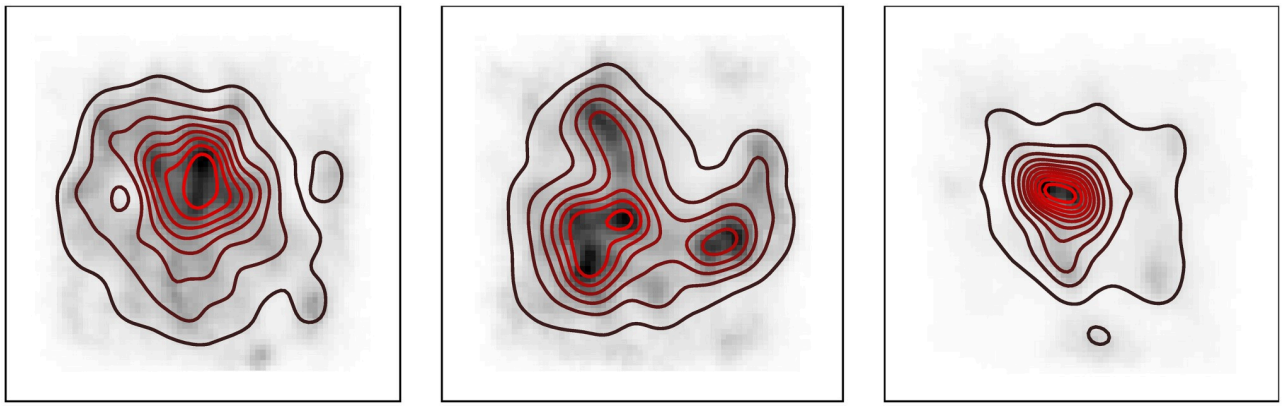

Fig 5. Comparison between the empirical saliency map and the fixation density of data generated by the model, for three different images from the test-set. The empirical saliency is represented by the shading, and the contour lines represent the density of the data generated by the model. The generated data recovers the original empirical saliency map.

https://doi.org/10.1371/journal.pcbi.1007880.g005

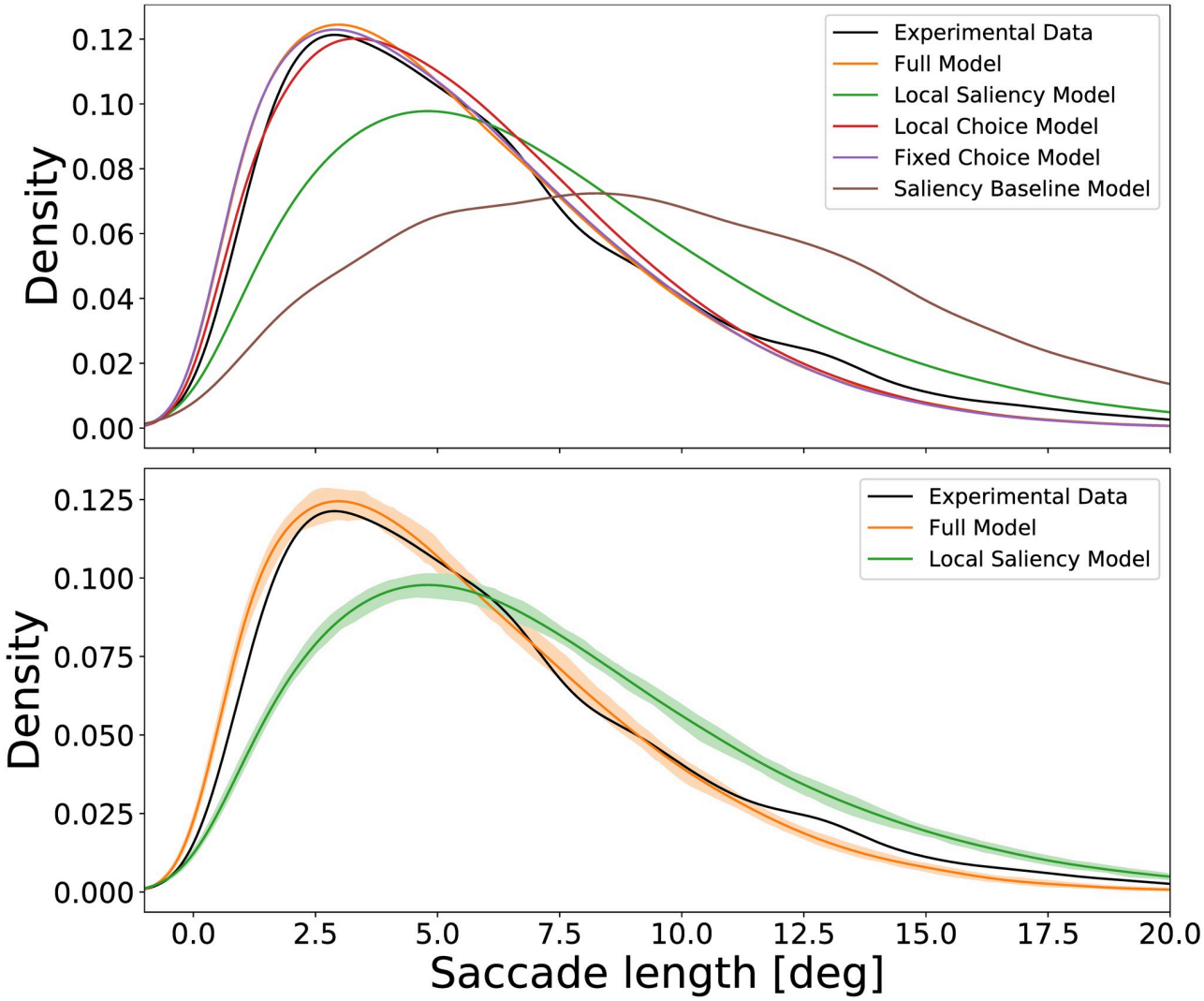

Fig 6. Saccade amplitude density—Experimental and simulated. Saccade amplitude density, aggregated over the data from all participants, of the experimental data and data generated by the full model and the simplified competitor models. *Top*. Comparison of all models. *Bottom*. Comparison between the full model and the Local Saliency Model. The shading corresponds to confidence bounds regarding the estimate of the model parameters. The full model captures the different kinds of saccade lengths, whereas the simpler models fail to do so.

https://doi.org/10.1371/journal.pcbi.1007880.g006

of subjects. The black curve presents the empirical data. The orange curve corresponds to data generated by the full Local and Global Attention Model, and the other curves correspond to the different simplified models.

As a baseline we include the Saliency Baseline Model, where the scanpath is sampled from the saliency map. As this model does not have any constraints on the saccade amplitude, other than the distance between high saliency areas in the image, the generated saccades have a much higher amplitude than the experimental data. Limiting the saccade amplitude as in the Local Saliency model by assuming a local attentional focus, results in saccades with much more realistic amplitudes. Fig 6 shows that the full model performs better than the simplified models. The three simplified models tend to capture the mean saccade amplitude rather than the full variety of saccade amplitudes displayed in scan paths. This behavior is expected from the Local Saliency Model, which includes only one type of characteristic saccade amplitude

whereas the full Local and Global Attention Model has two characteristic saccade amplitudes that correspond either to the local or global attention mode.

The Bayesian inference process presented in Methods Section results in a distribution over the possible values of the model parameters. This corresponds to uncertainty regarding the values of the model parameters. The shading around the generated data curves in Fig 6 corresponds to this uncertainty. We sampled 50 different values from the posterior distribution of each one of the model parameters and used this configuration to generate one data set. We split the experimental data into training and test sets three times and repeated the fitting of the models on each training set separately, resulting in a 5-fold cross-validation. This process applies to all the results presented, unless stated otherwise. Fig 6 presents the result of one such training and test split. The shading represents the 95% intervals around the mean density over the different data sets.

The confidence bounds are rather narrow and the density distributions of the two models are highly separable. This is a good indication that the Bayesian parameter inference is reasonable—the saccade amplitude density does not change dramatically with the parameter configurations sampled from the posterior distributions. The confidence bounds for the Local and Fixed Choice Models behave in a similar way and are not included in the figure for clarity purposes.

The model presented in this work generates a sequence of saccades, rather than independent saccades. Thus, we would expect to see some correlation between the generated saccades. Fig 7 presents the mean aurocorrelation of the saccade amplitude along a scan path. The experimental data shows a clear anti-correlation between the amplitude of subsequent saccades at lag 1. Thus, a short saccade is likely to be followed by a long saccade and vice-versa. Although

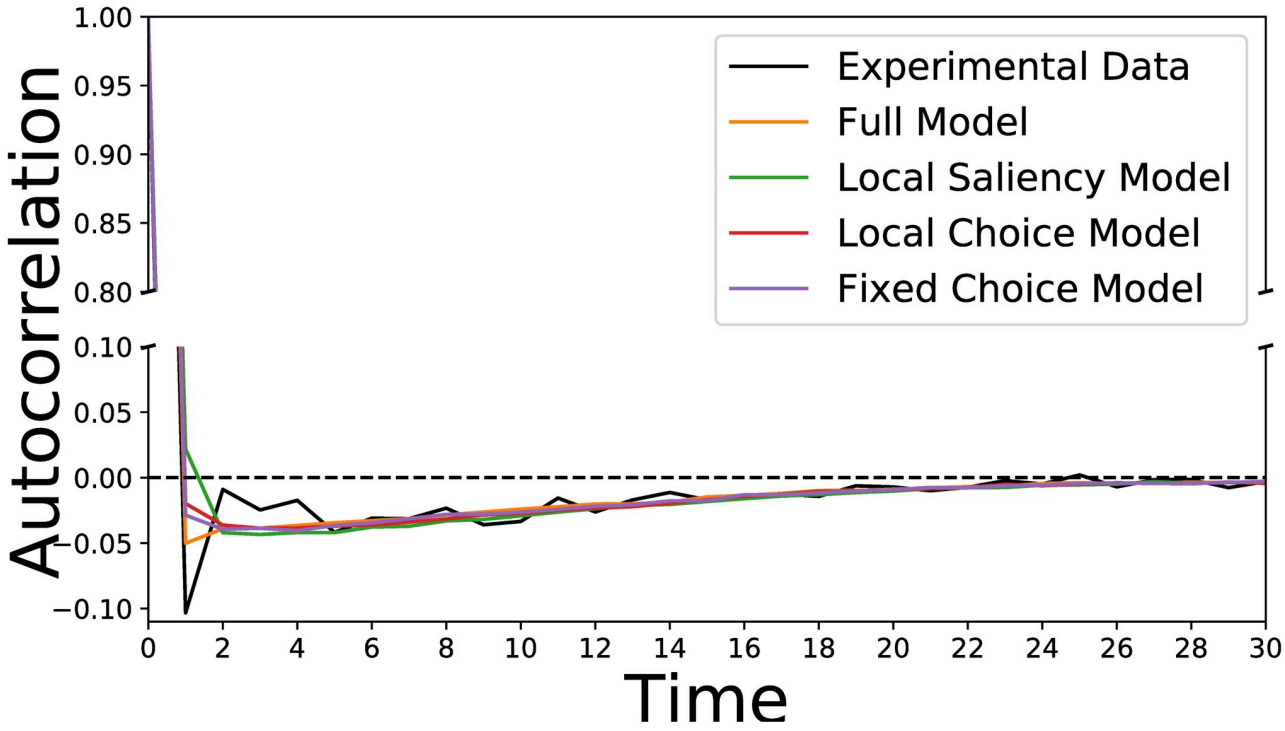

Fig 7. Saccade amplitude autocorrelation—Experimental and simulated. Saccade amplitude autocorrelation, averaged over experimental data from all participants and over all simulations generated by the full model (and the various competitor models). The full Local and Global Attention Model approximates the autocorrelation in amplitude of successive saccades, whereas the simpler models fail to reproduce the lag-1 anti-correlation.

https://doi.org/10.1371/journal.pcbi.1007880.g007

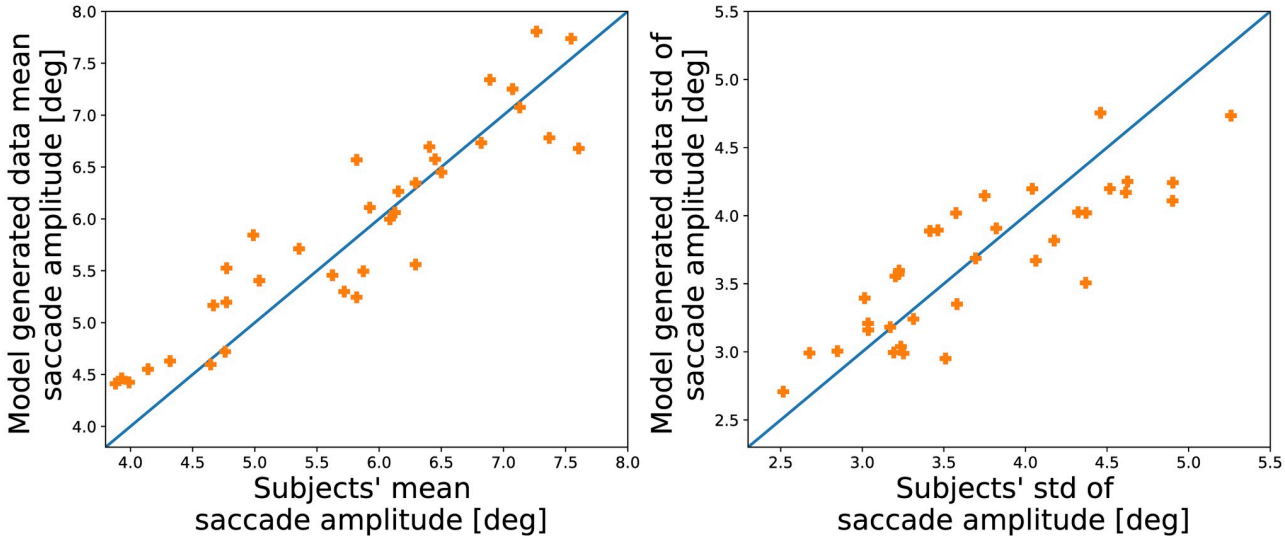

Fig 8. Comparison between experimental and simulated subjects' mean and standard deviation of saccade amplitude. *Left*. Participants' mean saccade amplitudes compared with the mean saccade lengths of data generated by the Local and Global Attention model. *Right*. The standard deviation of the subjects' saccade amplitude compared to the standard deviation of the data generated by the Local and Global Attention model. Overall the model captures the both the mean and the standard deviation of the saccade amplitude of the different subjects.

https://doi.org/10.1371/journal.pcbi.1007880.g008

not as strong as in the experimental data, this effect is captured by the Local and Global Attention Model. This result is expected from our modeling assumptions, since when generating fixation z_t the full model has information regarding the saliency of fixation z_{t-2}, whereas the competing models do not have access to this information. In addition to this lag-1 effect, it is important to note that our model also approximates the autocorrelation function for lags up to 20.

As described above, we fit a model for each subject individually. Thus, we can investigate how well the Local and Global Attention Model captures the difference between the subjects. In the left panel in Fig 8 we compare the mean saccade length of the empirical data and data generated from the full model for each subject. Each data point is one subject and the diagonal curve is the identity line. The presented data is from one fold of the k-fold cross validation.

Overall, the model captures the different mean saccade length of the different subjects. Not only does the model capture the different mean saccade amplitudes of the subjects, it also captures the difference in the variability of saccade amplitudes of the subjects (see the right panel of Fig 8, where the standard deviations of the saccade amplitudes are plotted per participant).

In Table 1 we report the coefficient of determination between the mean and standard deviation of the subjects' data and of the data generated by the full Local and Global Attention Model and the competing simplified model. The coefficient of determination was averaged across the different train-test splits in the cross validation. Other than the Local Saliency model, all model perform similarly well and capture the mean and standard deviation of the

Table 1. Comparison of the coefficient of determination, between the mean (or std) of the subjects' saccade amplitudes and the saccade amplitudes of the data generated by the different models. Other than the local saliency model, all models capture both the mean ant the standard deviation of the saccade amplitudes of the different subjects.

	Local and Global Attention	Local Choice	Fixed Choice	Local Saliency
R^2 saccade amplitude mean	0.93	0.93	0.93	0.47
R^2 saccade amplitude std	0.85	0.847	0.85	0.3

https://doi.org/10.1371/journal.pcbi.1007880.t001

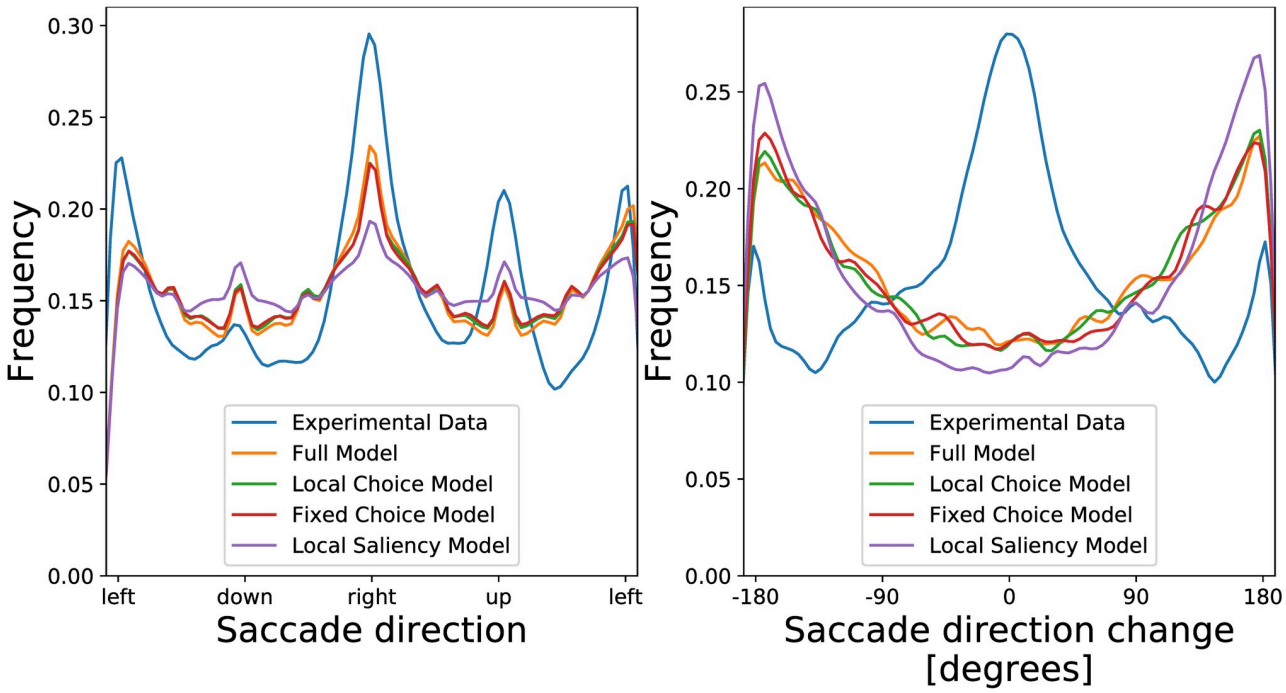

Fig 9. Experimental and simulated saccade direction and saccade direction change frequency. *Left*. Absolute saccade direction. The empirical data demonstrate a strong tendency to saccades towards the left and right directions, and a weaker tendency to perform saccades directed upwards. The generated data captures the tendency to perform horizontal saccades, but not vertical saccades. *Right*. Change in saccade direction. The generated data demonstrates the tendency to persist in the same direction. The models fail to capture this persistence.

https://doi.org/10.1371/journal.pcbi.1007880.g009

saccade amplitudes of the different subject. This result indicates that the assumption of two length scales generated by local and global attention states represents a major improvement in the model fit.

Saccade direction. Generally, saccades can be seen as vectors characterized by amplitude and direction. After analyzing the model performance with regard to the saccade amplitude, we turn to analyze the model performance with respect to the saccade direction.

There are two important aspects regarding saccade direction, i.e., absolute saccade direction and the direction relative to the previous saccade. In the left panel in Fig 9 we compare the saccade direction density, over the entire population of subjects, of the empirical data and of data generated by the fitted full model and its variations. The empirical data demonstrate clear preference for horizontal saccades and a weaker tendency towards vertical upward saccades. The different variations of the model generate similar distributions of saccade directions. The data generated by the models correspond to a tendency to perform horizontal saccades, but this tendency is not as strong as in the empirical data. The empirical tendency towards vertical saccades is not captured at all by the models.

The fact that the different models capture only one preferred horizontal direction is not surprising. It is common across all the variations of the model that at each step the next fixation is generated from an ellipsoidal distribution, which has only one preferred direction. In the Discussion, we suggest variations of the model which could capture more than one dominant saccade direction.

The right panel in Fig 9 presents the frequency of the values of the change in the saccade direction. The experimental data is characterized by a large peak around 0 which is an indication of persistence of the current saccade direction [35–37], also known as saccadic

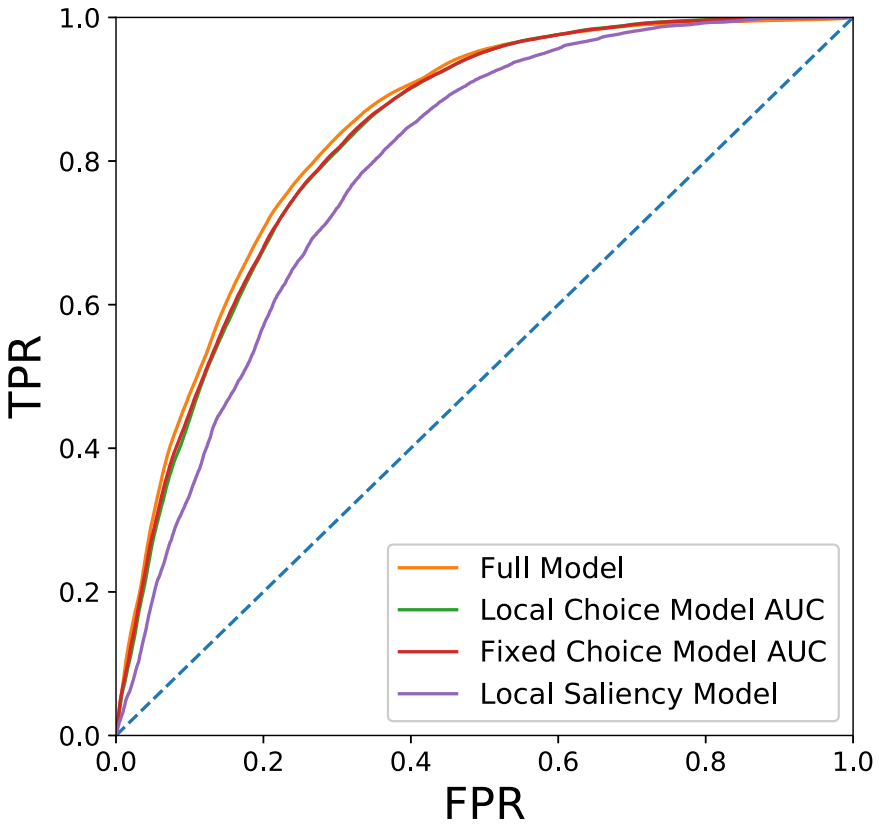

Fig 10. ROC curves of the different model variants. The full model performs slightly better than the Local and Fixed Choice models. The Local Saliency Model performs significantly worse than the other model variants.

https://doi.org/10.1371/journal.pcbi.1007880.g010

momentum. Additionally, there is a weaker peak around 180 and −180 degrees which indicates a tendency to return to the previous fixated location. All models discussed here fail to reproduce this effect. The peak in the saccade direction change is only around 180 and −180 degrees which is due to the hard constraints given by the image boundaries.

Model comparison. Last, we would like to compare the performance of the full Local and Global Attention Model to the simplified variants of the model presented in the Methods Section. As measurements of the model performance we use the Receiver Operating Characteristic (ROC), the respective Area Under the Curve (AUC), the Normalized Scan path Saliency (NSS) and the Information Gain (IG). The first two methods are widely used in the field of attention modeling and have been successfully adapted to scan path modeling [12, 14, 38–41]. The IG for saliency and scan path modeling analysis was first suggested by Kümmerer et al. in [42] and is becoming more common ever since [20, 43].

The AUC measure is a very common tool to analyze the performance of a probabilistic classifier by looking at the trade-off between True Positive and False Positive Rates (TPR and FPR), when using different thresholds for classification. In the context of attention modeling, the fixation locations are used as samples with positive labels, and as samples with negative label we used image coordinates that were sampled uniformly in the image space. This method is denoted as AUC-Borji [44]. In our analysis, the likelihood of the model at each time step is used as the probabilistic classifier. Fig 10 presents the ROC curves for the different models. We will discuss these results shortly after presenting the NSS measure.

Table 2. AUC, NSS and IG of the different model variants. The full model performs better than the Local and Fixed Choice models. The Local Saliency Model performs significantly worse than the other model variants.

	Local and Global Attention	Local Choice	Fixed Choice	Local Saliency
AUC	0.843	0.835	0.838	0.804
NSS	1.48	1.44	1.41	1.05
IG [bit/fix]	2.1	1.99	1.97	1.92

https://doi.org/10.1371/journal.pcbi.1007880.t002

The NSS is the average normalized saliency (with mean zero and standard deviation of one over the image) along the scan path of fixated locations. In this work we used the likelihood value of each fixation in the scan path, and normalized the likelihood over the entire image, as one would do with a saliency map. Table 2 presents the AUC and NSS scores for the different model variants.

Here, we use the same definition of IG as in [42]—as the average difference of the log-likelihood between a model and some baseline model. As a baseline we take a uniform distribution over the image, where the log-likelihood for each fixation is constant and equal to $\log_2$ (number of pixels). The units of this measure is bit per fixation (bit/fix) and it is interpreted as the amount of information gained in comparison to the baseline model, per fixation.

As done in the previous analysis, the scores are reported on the test data-set, which was not used in fitting data, and averaged over the different samples form the posterior distributions and the folds of the cross validation process. Due to the split of the data into training and test sets, the discussed measures are sensitive to the model complexity. If a model is too complex (usually manifested by having a lot of parameters) the model will achieve high AUC, NSS or IG, over the training set but may suffer from over fitting and perform poorly on the test set.

From the results in Table 2 we see that the full model performs better than the other variants, and achieves higher scores. Although the scores of the full model are the highest, they are only slightly better than the ones of the Local and Fixed Choice models. It is clear that the Local Saliency model performs poorly, which emphasizes the importance of the two characteristic length scales, which are present in all of the model variants other than the Local Saliency Model.

Discussion

The current study proposed and analyzed a mathematical model of fixation selection, motivated by the Local and Global Attention modes that were suggested previously as a mechanism driving eye movements in natural scene-viewing tasks [15, 21–23]. We constructed a generative scan path model based on a small set of assumption. Using Bayesian [45] inference we fit the model to experimental data. By doing so, we continue the line of work of using generative likelihood based models for scan path generation [20]. Importantly, we use recent developments in Bayesian statistics to construct more efficient parameter inference algorithms.

A different approach uses deep neural networks for scan path modeling [13, 14]. One of the downsides of this approach is its reliance on large amounts of data, which precludes the study of interindividual differences. Thus, by using a hypothesis–based model, which requires only a relatively small number of parameters, we can fit individual models for each experimental subject and capture inter–subject variability.

We demonstrate how our model captures the saccade amplitude both at the population and the individual level. Whereas two of the competing models perform equally well in terms of the coefficient of determination fitted to the mean and standard deviation of the individual subjects' saccade lengths, the advantage of the full model is demonstrated when looking at the

autocorrelation of the saccade length. This analysis takes into account not only the individual saccades but also the dynamics of the entire scan path. These results emphasize the importance of the information about the saliency of the previously fixated location, when deciding on the next fixation location, as described in Eq (7). Our model generates the typical behavior of a short saccade after a long one, or vice versa, which results in the observed anti-correlation of the length of subsequent saccades.

To further quantify the model performance we calculated the AUC, NSS and IG scores of the different variants of the model. The full model and the Local and Fixed Choice variants performed similarly well, and the full model achieved slightly higher scores than the other two. This result indicates that these quantitative scores may not be enough to evaluate how well a scan path model fits the data. Rather than relying only on likelihood based measures such as AUC, NSS and IG, a more careful investigation of the full body of results suggests that the saccade behavior is specifically characterized by its autocorrelation function of the saccades amplitudes.

Next, we will discuss the limitations of the model. As described above, the Local and Global Attention Model successfully captures the experimental saccade amplitudes both at the population level and the subject level. Another spatial aspect of saccades is saccade direction. Our model captures only the tendency to perform horizontal saccades, but not the tendency to perform vertical saccades. This is expected from the construction of the model.

As the full model has information regarding the saliency of the previous fixation location, but not of the location itself, in its current form the model does not capture the change in saccade direction (i.e., the saccade direction relative to the previous saccade). The relative saccade direction is important for modeling known phenomena such as visual persistence or saccadic momentum [35–37, 46]. As with the vertical preferred saccade direction, the model's inability to capture the relative saccade direction stems from the choice of Gaussian functions in the local and global attentional states. Our model could be extended to account for these tendencies by a mixture of Gaussians. Each Gaussian component would be designed to capture different directional tendencies, rather than capturing only one tendency as in the current version of the model. For example, one Gaussian can be aligned in the direction of the previous saccade, to account for visual persistence.

Other limitations of the model stem from the choice of a second order Markov process. Due to this choice the model is almost memory-less and cannot capture known phenomena in scene viewing which span multiple saccades. Incorporating longer history is not straightforward in our model. A heuristic approach could be including dynamics in the saliency map.

Finally, our mathematical model does not account for fixation duration in scene viewing, which of course play an important role in eye-movement control [19, 47–49]. So far most of the modeling attempts of scene viewing addressed either the spatial or the temporal aspects of scene viewing. Indeed, some models use temporal dynamics but they do not attempt to learn these dynamics from the data and use a heuristic–based approach. While fixation duration modeling is outside the scope of this work, we nonetheless consider the integration of temporal and spatial aspects an exciting new research direction.

Supporting information

S1 Text. Parameter inference. In this appendix you can find the technical details of the Gibbs sampler described in the Methods section. It includes a derivation of the full likelihood, details regarding the augmentation schemes and the derivation of the conditional distributions.
(PDF)

Author Contributions

Conceptualization: Noa Malem-Shinitski, Manfred Opper, Ralf Engbert.

Data curation: Ralf Engbert.

Formal analysis: Noa Malem-Shinitski, Manfred Opper.

Funding acquisition: Sebastian Reich.

Investigation: Noa Malem-Shinitski.

Methodology: Noa Malem-Shinitski, Manfred Opper.

Resources: Sebastian Reich, Ralf Engbert.

Software: Noa Malem-Shinitski, Lisa Schwetlick, Stefan A. Seelig.

Supervision: Manfred Opper, Sebastian Reich, Ralf Engbert.

Validation: Noa Malem-Shinitski, Ralf Engbert.

Visualization: Lisa Schwetlick.

Writing – original draft: Noa Malem-Shinitski, Ralf Engbert.

Writing – review & editing: Noa Malem-Shinitski, Sebastian Reich, Ralf Engbert.

References

1. Chalupa LM, Werner JS. The Visual Neurosciences, Vols. 1 & 2. MIT Press; 2004.
2. Findlay JM, Gilchrist ID. Active Vision: The Psychology of Looking and Seeing. Oxford University Press; 2003.
3. Gameiro RR, Kaspar K, König S, Nordholt S, König P. Exploration and Exploitation in Natural Viewing Behavior. Scientific Reports. 2017; 7(1):1–23.
4. Ehinger BV, Kaufhold L, König P. Probing the temporal dynamics of the exploration–exploitation dilemma of eye movements. Journal of Vision. 2018; 18(3):6–6. https://doi.org/10.1167/18.3.6
5. Berger-Tal O, Nathan J, Meron E, Saltz D. The exploration-exploitation dilemma: a multidisciplinary framework. PloS one. 2014; 9(4). https://doi.org/10.1371/journal.pone.0095693 PMID: 24756026
6. Bisley JW, Mirpour K. The neural instantiation of a priority map. Current Opinion in Psychology. 2019; p. 108–112. https://doi.org/10.1016/j.copsyc.2019.01.002
7. Itti L, Koch C. A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Research. 2000; 40(10-12):1489–1506. https://doi.org/10.1016/S0042-6989(99)00163-7
8. Borji A, Itti L. State-of-the-art in visual attention modeling. IEEE Transactions On Pattern Analysis And Machine Intelligence. 2012; 35(1):185–207. https://doi.org/10.1109/TPAMI.2012.89
9. Kümmerer, Theis, Bethge. Deep Gaze I: Boosting saliency prediction with feature maps trained on imagenet. Preprint arXiv:14111045. 2014;.
10. Kümmerer, Wallis, Bethge. Deep Gaze II: Reading fixations from deep features trained on object recognition. Preprint arXiv:161001563. 2016;.
11. Einhäuser W, König P. Getting real—sensory processing of natural stimuli. Current Opinion in Neurobiology. 2010; 20(3):389–395. https://doi.org/10.1016/j.conb.2010.03.010
12. Kümmerer M, Wallis TSA, Bethge M. Saliency Benchmarking Made Easy: Separating Models, Maps and Metrics. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y, editors. Computer Vision—ECCV 2018. Lecture Notes in Computer Science. Springer International Publishing; 2018. p. 798–814.
13. Shao X, Luo Y, Zhu D, Li S, Itti L, Lu J. Scanpath prediction based on high-level features and memory bias. In: International Conference on Neural Information Processing. Springer; 2017. p. 3–13.
14. Kümmerer M, Wallis TS, Bethge M. DeepGaze III: Using Deep Learning to Probe Interactions Between Scene Content and Scanpath History in Fixation Selection. In: 2019 Conference on Cognitive Computational Neuroscience, 13-16 September 2019, Berlin, Germany; 2019.
15. Tatler BW, Vincent BT. Systematic tendencies in scene viewing. Journal of Eye Movement Research. 2008; 13(12):1–18.

16. Zelinsky GJ. A theory of eye movements during target acquisition. Psychological review. 2008; 115 (4):787–835 https://doi.org/10.1037/a0013118 PMID: 18954205
17. Le Meur O, Liu Z. Saccadic model of eye movements for free-viewing condition. Vision Research. 2015; 116:152–164. https://doi.org/10.1016/j.visres.2014.12.026
18. Engbert R, Trukenbrod HA, Barthelmé S, Wichmann FA. Spatial statistics and attentional dynamics in scene viewing. Journal of Vision. 2015; 15(1):14–14. https://doi.org/10.1167/15.1.14
19. Tatler BW, Brockmole JR, Carpenter RH. LATEST: A model of saccadic decisions in space and time. Psychological Review. 2017; 124(3):267–300 https://doi.org/10.1037/rev0000054 PMID: 28358564
20. Schütt HH, Rothkegel LO, Trukenbrod HA, Reich S, Wichmann FA, Engbert R. Likelihood-based Parameter Estimation and Comparison of Dynamical Cognitive Models. Psychological Review. 2017; 124(4):505. https://doi.org/10.1037/rev0000068
21. Frost D, Pöppel E. Different programming modes of human saccadic eye movements as a function of stimulus eccentricity: Indications of a functional subdivision of the visual field. Biological Cybernetics. 1976; 23(1):39–48.
22. Unema PJ, Pannasch S, Joos M, Velichkovsky BM. Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. Visual Cognition. 2005; 12(3):473–494.
23. Helmert JR, Joos M, Pannasch S, Velichkovsky BM. Two visual systems and their eye movements: Evidence from static and dynamic scene perception. In: Proceedings of the Annual Meeting of the Cognitive Science Society; 2005. p. 2283–2288.
24. Cohen JD, McClure SM, Yu AJ. Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. Philosophical Transactions of the Royal Society B: Biological Sciences. 2007; 362(1481):933–942. https://doi.org/10.1098/rstb.2007.2098
25. Berger-Tal O, Nathan J, Meron E, Saltz D. The exploration-exploitation dilemma: a multidisciplinary framework. PloS One. 2014; 9(4):e95693. https://doi.org/10.1371/journal.pone.0095693
26. Tatler BW, Baddeley RJ, Vincent BT. The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task. Vision Research. 2006; 46(12):1857–1862. https://doi.org/10.1016/j.visres.2005.12.005
27. Geman S, Geman D. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. In: Readings in Computer Vision. Elsevier; 1987. p. 564–584.
28. Liu JS. Monte Carlo strategies in scientific computing. Springer Science & Business Media; 2008.
29. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian data analysis. Chapman and Hall/CRC; 2013.
30. Polson NG, Scott JG, Windle J. Bayesian Inference for Logistic Models Using Pólya–Gamma Latent Variables. Journal of the American Statistical Association. 2013; 108(504):1339–1349. https://doi.org/10.1080/01621459.2013.829001
31. Wong A. Eye movements; saccades. Encyclopedia of the neurological sciences; 2014.
32. Gilks WR, Wild P. Adaptive Rejection Sampling for Gibbs Sampling. Journal of the Royal Statistical Society: Series C (Applied Statistics). 1992; 41(2):337–348.
33. Martino L, Yang H, Luengo D, Kanniainen J, Corander J. A fast universal self-tuned sampler within Gibbs sampling. Digital Signal Processing. 2015; 47:68–83. https://doi.org/10.1016/j.dsp.2015.04.005
34. Duane S, Kennedy AD, Pendleton BJ, Roweth D. Hybrid Monte Carlo. Physics Letters B. 1987; 195 (2):216–222. https://doi.org/10.1016/0370-2693(87)91197-X
35. Ritter M. Evidence for visual persistence during saccadic eye movements. Psychological Research. 1976; 39(1):67–85. https://doi.org/10.1007/BF00308946
36. Breitmeyer BG, Kropfl W, Julesz B. The existence and role of retinotopic and spatiotopic forms of visual persistence. Acta psychologica. 1982; 52(3):175–196. https://doi.org/10.1016/0001-6918(82)90007-5
37. Wilming N, Harst S, Schmidt N, König P. Saccadic momentum and facilitation of return saccades contribute to an optimal foraging strategy. PLoS Computational Biology. 2013; 9(1):e1002871. https://doi.org/10.1371/journal.pcbi.1002871
38. Peters RJ, Iyer A, Itti L, Koch C. Components of bottom-up gaze allocation in natural images. Vision Research. 2005; 45(18):2397–2416. https://doi.org/10.1016/j.visres.2005.03.019
39. Wang W, Chen C, Wang Y, Jiang T, Fang F, Yao Y. Simulating human saccadic scanpaths on natural images. In: CVPR 2011. IEEE; 2011. p. 441–448.
40. Borji A, Itti L. State-of-the-art in visual attention modeling. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2012; 35(1):185–207. https://doi.org/10.1109/TPAMI.2012.89

41. Riche N, Duvinage M, Mancas M, Gosselin B, Dutoit T. Saliency and human fixations: State-of-the-art and study of comparison metrics. In: Proceedings of the IEEE international conference on computer vision; 2013. p. 1153–1160.

42. Kümmerer M, Wallis TS, Bethge M. Information-theoretic model comparison unifies saliency metrics. Proceedings of the National Academy of Sciences. 2015; 112(52):16054–16059. https://doi.org/10.1073/pnas.1510393112

43. Schwetlick L, Rothkegel L, Trukenbrod H, Engbert R. Modeling the effects of perisaccadic attention on gaze statistics during scene viewing. Preprint https://doi.org/10.31234/osf.io/zcbny.

44. Bylinskii Z, Judd T, Oliva A, Torralba A, Durand F. What do different evaluation metrics tell us about saliency models? IEEE Transactions on Pattern Analysis and Machine Intelligence. 2018; 41(3):740–757. https://doi.org/10.1109/TPAMI.2018.2815601

45. MacKay DJ, Mac Kay DJ. Information theory, inference and learning algorithms. Cambridge university press; 2003.

46. Luke SG, Smith TJ, Schmidt J, Henderson JM. Dissociating temporal inhibition of return and saccadic momentum across multiple eye-movement tasks. Journal of Vision. 2014; 14(14):9–9. https://doi.org/10.1167/14.14.9

47. Henderson JM. Human gaze control during real-world scene perception. Trends in Cognitive Sciences. 2003; 7(11):498–504. https://doi.org/10.1016/j.tics.2003.09.006

48. Nuthmann A, Smith TJ, Engbert R, Henderson JM. CRISP: a computational model of fixation durations in scene viewing. Psychological Review. 2010; 117(2):382. https://doi.org/10.1037/a0018924

49. Laubrock J, Cajar A, Engbert R. Control of fixation duration during scene viewing by interaction of foveal and peripheral processing. Journal of Vision. 2013; 13(12):11–11. https://doi.org/10.1167/13.12.11