Data Assimilation for Neurocognitive Models of Eye Movement

A Doctoral Dissertation
submitted to the
University of Potsdam

By
Lisa Schwetlick

Supervisor & First Corrector: Prof. Dr. Ralf Engbert
Second Corrector: Prof. Dr. Casimir Ludwig
Date of Submission: 23.03.2023, Potsdam

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-59828
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-598280

Abstract

Visual perception is a complex and dynamic process that plays a crucial role in how we perceive and interact with the world. The eyes move in a sequence of saccades and fixations, actively modulating perception by moving different parts of the visual world into focus. Eye movement behavior can therefore offer rich insights into the underlying cognitive mechanisms and decision processes. Computational models in combination with a rigorous statistical framework are critical for advancing our understanding in this field, facilitating the testing of theory-driven predictions and accounting for observed data. In this thesis, I investigate eye movement behavior through the development of two mechanistic, generative, and theory-driven models. The first model is based on experimental research regarding the distribution of attention, particularly around the time of a saccade, and explains statistical characteristics of scan paths. The second model implements a self-avoiding random walk within a confining potential to represent the microscopic fixational drift, which is present even while the eye is at rest, and investigates the relationship to microsaccades. Both models are implemented in a likelihood-based framework, which supports the use of data assimilation methods to perform Bayesian parameter inference at the level of individual participants, analyses of the marginal posteriors of the interpretable parameters, model comparisons, and posterior predictive checks. The application of these methods enables a thorough investigation of individual variability in the space of parameters. Results show that dynamical modeling and the data assimilation framework are highly suitable for eye movement research and, more generally, for cognitive modeling.

Acknowledgments

First and foremost I would like to thank Ralf Engbert, my supervisor for all the help, guidance, and discussions, and for believing in the solution when I was close to giving up. I also thank my colleagues for the many interesting and constructive conversations and support. Particularly Daniel, for being the best office companion I could wish for, and Noa, for discussing models with me and being there for me along the way. A big thank you goes to Petra Schienmann and everyone who worked in the Eye Lab, collecting data, coordinating with study participants and annotating data: neither this thesis, nor any of the work in our lab would be possible without you!

This PhD has been a journey with all the ups and downs that might be expected. I am deeply indebted to the people in my life that inspired, supported, and encouraged me personally, particularly my family, Clara and Susanne, and Jake. Thanks also to all my friends who graciously endured my confused ramblings and brought me back to the plane of reality when I needed it.

Lastly, I dedicate these pages to my dad. Thank you, for sharing your love of science and your enthusiasm; I wish you could be here to discuss it still.

Contents

1 Introduction

All models are wrong but some are useful

G. E. P. Box

Molecules diffuse along a chemical gradient; a neuron transmits an electrical charge to its neighbor; cells connect to form a network structure; and somehow from this process emerges cognition, creativity, emotion, and behavior. Understanding information processing in the complex system that is the brain has been a major scientific drive in the last centuries. The question of how humans and other intelligent animals produce complex behaviors has fascinated scientists and thinkers. Modern cognitive- and neurosciences seek to find the relations between behavior, cognition, neural activity in the nervous system, and the environment.

Complex cognition and behavior rely on a continual interaction between perception and action. This is particularly apparent in the case of visual perception. Eye movements shift the visual input over the receptors in the eye, bringing different parts of the visual world into the center of focus. These movements occur at both macroscopic and microscopic levels, such that the signal that enters the eye is constantly changing. The decision processes that underlie eye movement behavior are determined by both high-level cognition and low-level features and are the result of a constant integration of perception and action. This interplay is a core component of visual perception, not an inconvenient necessity that has to be accounted for in the processing of the signal. It is in good agreement with the idea that the architecture of the brain responds to changes in signal rather than to static signals. At the level of microscopic eye movements, shifts are related to preventing response fatigue of receptors and improving visual acuity. Macroscopic eye movements offer insight into cognitive decision processes, perception, and action planning.

As in many other scientific disciplines, models are of central importance for advancing our understanding of the brain. A model is an informative representation of a system that may be derived bottom-up from data, or top-down from theory. The first case subsumes much of statistical modeling, where models describe the relationship of observable variables. Theory-driven models, on the other hand, represent hypotheses by formulating laws and axioms in a normative way and relating them to observable data. In the case of computational models this typically means a concrete implemen-

tations which may depart from, extend, or refine the theory in order to accommodate the implementation requirements and aims. Computational modeling is powerful tool to understand which predictions arise from theories and to which extent the current state of knowledge can account for the observed data.

Within this thesis, I investigate eye movement behavior by developing two mechanistic, theory-driven models. I begin the thesis with an introduction to the field of scientific modeling and the applied modeling framework. In the following chapter I introduce some of the core concepts of eye movement research, including different types of movement and existing models. The introduction is followed by the three examples of eye movement modeling.

First, I develop a model of scan paths, which implements findings about the distribution of visual attention to show that statistical characteristics of eye movement trajectories can be explained by first order principles. Specifically, I show that a functional redistribution of attention around the time of the saccade has an impact on the fixation selection process (Schwetlick, Rothkegel, Trukenbrod, et al., 2020a). The presented implementation is likelihood-based and allows Bayesian statistical inference of parameters at the level of individual subjects. These results are presented in Chapter 2. In Chapter 3, I further extend the model to make predictions of fixation durations in addition to fixation locations by introducing a spatiotemporal likelihood function. I then apply this extended SceneWalk model to the question of how different tasks influence the choice of fixation locations (Schwetlick, Backhaus, & Engbert, 2022a).

The second model, presented in Chapter 4, focuses on microscopic eye movement and proposes a relationship between different types of movement. Specifically, the model is a self avoiding random walk within a confining potential, which represents slow, meandering fixational drift. I show that the fitted parameters of this likelihood-based model are distinct for individual subjects. Additionally, I use the latent model parameters to investigate the interaction of fixational drift and microsaccades.

Chapter 5 discusses the results, the merits of the applied modeling framework and provides an overarching discussion. The results show that dynamical modeling is highly suitable for the field of eye movement research. Specifically in all three examples it is possible to fit parameters to a model for individual subjects. Furthermore, the parameters of the model are interpretable and allow a thorough investigation of the individual variability in the space of parameters.

1.1 Modeling as a tool for understanding the brain

Real-world systems often depend on highly complex mechanisms and interactions; the brain is a *model*[1] exemplar of this. Studying such a system directly may only

1 The semantics of the word *model* depend on the context and the field. Usage ranges from the scaled down version of a building that an architect might make; to comparing atoms to oranges in first year chemistry; to a person showing off clothes for a brand; to the computational implementation of a complicated scientific theory. Their commonality is that, on one level or another, they are

be possible under certain limitations and the observable data may be multi-causally determined and difficult to interpret. Scientific models are a way of studying complex systems indirectly, usually by using the model to predict outcomes and compare them to empirical data. In the following paragraphs I will outline the value of models, predictions, and their interaction with empirical data.

First, in a complex system observable data is often also complex and influenced by many factors and mechanisms. Explaining the relationships between observed data and determining causes and effects is a core objective of scientific research and also an essential component of model-building. However, in the field of psychology and cognitive science, laws and theories in the traditional sense are rare. Instead, findings are usually thought of as effects that tend to characterize phenomena rather than explain them (Bechtel & Abrahamsen, 2010; Cummins, 2010). When ideas and hypotheses about underlying mechanisms are put forward, they are typically not formalized. However, formalizing a model is an important step in building a functional understanding of the system and considering the implications of the proposed mechanism. An algorithmic or mathematical description, while not a guarantee of scientific rigour, requires concrete details that can be glossed over in verbal descriptions and provides more tools to evaluate its value (Bechtel & Abrahamsen, 2005). Moreover, when a theory exists, it is only through models that interpret these concepts concretely that it may be tested (Cartwright, 1999).

The predictions of models are also valuable as strong tests of the underlying theories. When the mechanism underlying the phenomenon can be adequately described by a model, the predicted outcomes should match the empirically measured outcomes. Any observed deviance from the prediction can be interpreted as a part of the phenomenon that is not sufficiently understood or implemented. Predicted and true outcomes may be compared using posterior predictive checks, as outlined in Section 1.2.4. When a model can explain (parts of) data, it gives credibility to the proposed mechanism. (Bechtel & Abrahamsen, 2010). On the other hand, failure to predict certain effects or edge cases, may inspire ideas about how to improve the model and scientific understanding.

When many mechanisms interact, their behavior may be different than the sum of their parts. Complex interactions are not always easy or possible to predict. Such emergent behaviors can be important for understanding the behavior of a system. When a model exhibits a behavior that was not explicitly built-in, but which nonetheless aligns with the data, this may be considered a strong confirmation about the proposed mechanisms. In contrast, it is also possible, using a model, to identify which of the proposed parts and mechanisms are necessary and which superfluous to produce the target behavior, by constructing model versions with and without certain components (see Bechtel and Abrahamsen, 2010 and for another application of this approach Schmittwilken and Maertens, 2022).

Furthermore, models are an opportunity to study systems under manipulations that

intended as representations of some real-world system.

are impractical or impossible to explore empirically. It is usually possible to change the parameters of a model much more easily than to setup a sufficiently large scale empirical study. In a similar vein, particularly in the life sciences where individual variation tends to be large, the response of a model given parameter differences, may be valuable in understanding the size and robustness of the observed effects.

Lastly, predictions of real-world phenomena have a practical real-world value, e.g., when you bring an umbrella after seeing the weather forecast. Such models may be used to predict earthquakes, assist drivers, or divine stock markets. In order to have such a practical purpose, application models must fulfill different criteria than purely scientific ones. Predictions need to be robust outside of idealized laboratory conditions, which may include a limited availability of data. In the modern world computational models are ubiquitous in almost every area of life, underlying everything from computer vision to production chains and advertisement. More relevant for the scientific context are models that may serve as a baseline for further experiments; a model of the reflective properties of the eye underlies the video-based eye tracking devices in our lab (SensoMotoric Instruments, 2016; YutaItoh, 2016); complicated models of conductivity underlie the analysis of every EEG and MRI recording (Næss et al., 2021).

Ontologies of models are many and varied. Model-types may be classified by their description-level (e.g., verbal, mathematical, algorithmic, diagrammatic), by their structure (e.g., agent-based, dynamical, concrete, neural network), purpose (e.g., predictive, procedural), or their inspiration (e.g., theory-driven, data-driven). These (and other) dimensions are continuous and a model can fall anywhere along the spectrum (Gershenfeld, 1999). For the purposes of this thesis I will focus on models that can be expressed mathematically and implemented computationally. Specifically, I explore the advantages of theory-driven, dynamical mechanistic modeling in cognitive science.

1.1.1 Dynamical mechanistic modeling

Cognition is an active and dynamic process in a biological, neurophysiological system. An emphasis on these features, dynamics and biological plausibility, are fundamental to researching and understanding behavior. The first aspect reflects the fact that perception and internal processing, and action are not independent and they change dynamically over time (Van Gelder & Robert, 1995). The latter refers to the idea that biological and neurophysiological findings should provide the boundary conditions for behavioral explanations. Mechanistic models explain behavior at the level of how the component parts of a system work together to produce an effect (Bechtel & Abrahamsen, 2005). A model that can show that observed behavior emerges from biologically plausible, dynamical mechanisms is a strong test of the underlying assumptions.

A popular analogy in cognitive science is that the brain is like a computer, in that it receives input, processes information, and produces output behavior (Newell, Simon, et al., 1972). However, biological systems can not be fully described as input/output machines. Instead, as stated by Van Gelder and Robert (1995), "cognitive processes and their context unfold continuously and simultaneously in real time". The con-

sequence of taking this observation seriously is that efforts to understand cognition should also emphasize the dimension of time. Dynamical models are popular in other areas of science to describe everything from fluid flow to solar systems and have great potential in cognitive science. An early example of a dynamical model in cognitive science is the Haken, Kelso, Bunz model (1985), which models phase transitions in human finger movements using coupled non-linear oscillators. The range of dynamical models in cognitive science include models for movement preparation (Erlhagen & Schöner, 2002), decision-making (Busemeyer & Townsend, 1993) and sensorimotor integration (Churchland, 1989) (a collection of further models can be found in Port and van Gelder, 1995).

In a dynamical model every model state can be represented numerically and evolves over time according to some rule. The current state of the model depends uniquely on the rule and the previous states. The rule for the evolution over time (i.e., the rate of change) is usually stated as a (set of) differential equation(s) (Gershenfeld, 1999). A key feature of this architecture is that the model can be evaluated for any point in time to investigate the model states. When constructed mechanistically, the model states and parameters represent interpretable quantities. Thus, model predictions are available that may be compared to the corresponding empirical observations at the corresponding time. Dynamical models allow us to leverage the full information available in behavioral data, including dependencies in the data over time (Van Gelder & Robert, 1995).

The argument for mechanistic modeling follows the idea that observable behavior is produced and constrained by the underlying (biological) system. Researchers typically seek to to understand how and why a phenomenon is produced. An explanation that is firmly grounded in the properties of the component parts, has several advantages. It can show whether the proposed orchestration of components is capable of producing the observed behavior, allows exploration about which components are strictly necessary and how alterations to the structure change the outcome. Most importantly, mechanistic modeling is tightly coupled with experimental research: models can be built on the existing knowledge, investigate conclusions, and uncover aspects of the data that are not fully understood. An impressive case in point for mechanistic modeling is made by the field of circadian rhythm research (see Bechtel and Abrahamsen, 2010). Research of the circadian system is notable because modeling approaches have leveraged the mechanistic modeling framework in a remarkably effective way. In one example, experimental researchers carefully identified a negative feedback loop of how, in multiple steps the transcription of a gene influences concentrations of proteins which then slow the transcription. A model of this loop showed that it indeed leads to the kind of periodic behavior that would be expected from a circadian timekeeping mechanism (Goldbeter, 1995).

A key feature of cognition, perception, and action is that they unfold continuously over time within a biological system. Investigating these processes within their natural framework, instead of isolated from it, has proven to be a compelling line of research (Bechtel & Abrahamsen, 2010; Port & van Gelder, 1995). Dynamical, mechanistic modeling approaches are therefore of particular interest and usefulness for broadening

our scientific understanding of such processes. A second important advantage of dynamical, process oriented, biologically plausible models is that, as they are grounded in empirical evidence, specific assumptions can be tested against empirical data. Thus, within this framework is particularly well-suited as the basis for modeling behavior that is shaped by the underlying cognitive processes. In this thesis I use this approach to develop models of human eye movement.

1.1.2 Symbolic cognitive modeling

Aside from dynamical models, there are also other forms of procedural cognitive modeling. ACT-R (Adaptive Control of Thought - Rational), a symbolic model of higher level cognition, was originally developed by J. R. Anderson and Bower (1973). It uses explicit rules and symbols to represent and manipulate information, allowing for the simulation of high-level cognitive processes such as decision-making and problem-solving. ACT-R is based on the idea that the mind is composed of a number of "modules" that are specialized for different kinds of tasks, such as perception, memory, and decision-making. Each module has a set of rules or procedures that it follows to perform its task and these rules are activated by the contents of working memory (Ritter et al., 2018). Within its modules, ACT-R also uses connectionist approaches, for example to analyze input patterns in the perception modules or to store information in the memory modules. It is now widely recognized that a comprehensive model of cognition will likely require a combination of both symbolic and connectionist approaches.

ACT-R can be used to model a variety of different problems, although it is primarily a model of learning. This generality comes with advantages and disadvantages. On one hand, using a general framework allows results to be more easily interpreted and integrated within other results in the same context. On the other hand, in order to model something specific and new the general framework may need to be extended and adapted where otherwise it may be more straight-forward to focus on the core concepts in a stand-alone model. A dynamical model, when defined and optimized in a statistically rigorous way, is more parsimonious in terms of its structure and still confers most of the advantages like comparability and the ability to integrate into other systems, in addition to placing the emphasis on the specific features under investigation.

1.1.3 Data-driven modeling

A different approach to mathematical modeling prioritizes descriptiveness over procedural analysis. Data-driven modeling has surged in popularity as resources like computing power and immense data sets have become available (Maass et al., 2018). In contrast to theory-driven model development, which starts by developing a theoretical framework and then analyzes the data, data-driven research seeks to exploratively extract scientifically interesting insights from data (Kelling et al., 2009) usually without an underlying theory. A statistical analysis like linear regression, for example, clearly

falls into this category: two variables are related to each other, but typically without explicitly proposing any functional relationship. Advances in statistical methods allow complex, data-driven techniques including multivariate analyses, linear mixed effects modeling and fitting distributions, as well as neural network modeling.

Statistical modeling forms the basis for many experimental scientific disciplines. It encompasses a suite of tools to get numerical confirmation about the significance of study outcomes. A danger of modeling in this context, especially in the fields of psychology and cognitive science (see the replication crisis, Collaboration, 2015), is that standardized model recipes used without careful scrutiny of requirements may lead to unjustified conclusions. The blind application of statistical modeling and especially the heavy reliance on p-values has been criticized (Matthews, 2000). However, statistical modeling is by no means limited to the application of "Hypothesis Testing recipes" (Kaplan, 2009). Custom statistical models may rely on fitting distributions to data, exploring data using clustering or dimensionality reduction, and can be extraordinarily useful to understand the relationship between observable variables and to start building theories about the underlying structures (Engbert, 2021).

The difference to theory-driven models, like the dynamical, mechanistic models discussed in this thesis, is clear. It is more broadly applicable and often requires fewer custom procedures. On the other hand, data-driven modeling does not explicitly have an emphasis on understanding the processes that drive the observed behavior and relies on deliberate and strong experimental design to be able to make statements about the relationships between variables. The role of data-driven modeling can not be underestimated, particularly in the stage of approaching a phenomenon and developing hypotheses, as well as after model development, for concisely testing outcomes. However, a procedural approach is indispensable for gaining a deeper understanding about the underlying processes and investigating functional hypotheses.

1.1.4 Neural networks and connectionism

Another direction at the intersection of data-driven and procedural modeling is neural network modeling. Simply stated, neural networks are constructed by imposing a layered architecture of connected “Neurons”, where each connection applies some mathematical transformation. The connections are parameterized with weights and biases. Fitting a neural network involves finding appropriate values for all of these parameters using a loss function (Nielsen, 2015). This basic approach has been much developed recent years, due to the success of this method to perform complicated operations. Applications range from computer vision models that can determine the content of a scene (e.g., Simonyan & Zisserman, 2014), to language models that can generate strings of words that are often indistinguishable to human-produced sentences and texts (Brown et al., 2020). The impressive ability of neural network models to perform such tasks relies on immense amounts of computing time and data to find appropriate values for the many thousands of parameters necessary to represent the complexity of the problem. A caveat of this approach, particularly for brain research, is that while neural networks perform well as input/output machines, the analysis of

their internal states is far from trivial.

A popular aspect of neural network research is the intersection with cognitive science: already the term "neural network" places the approach in relation to neurons in the brain. It is tempting, therefore, to interpret neural networks as models of the brain, e.g., to find that neural networks detect similar features or develop similar activity patterns as measured by electrophysiology (e.g., Cadena et al., 2019). However, computational neural networks behave differently from biological neural networks in many ways (Funke et al., 2020; Geirhos et al., 2018). This becomes particularly apparent with respect to robustness to noise (Geirhos et al., 2017).

In Chapter 5, I discuss some preliminary results from using neural networks and dynamical, mechanistic models to compare and contrast their respective advantages and disadvantages. I suggest that a joint approach is mutually beneficial to both fields. Neural networks usually do not mirror biological properties of individual units, nor is it clear that artificial neural networks learns concepts in any way similar to biological neural networks. However, their highly versatile nature for solving a variety of problems makes them an interesting tool also for scientific use cases and it is valuable to explore this potential.

1.2 The modeling framework

The main part of this thesis is comprised of two different theory-based models of different levels of eye movement data. The presented modeling framework relies on 3 components: the model and its implementation, the parameters and their estimation, and posterior predictive checks. The first step in any modeling research project, designing and building an appropriate model, is an art with much freedom of choice. Even when formalized theories exist to back a model, implementation details tend to play a significant role (Cartwright, 1999). The model implementation represents the hypothesized structure or mechanisms which, in the ideal case, provide a parsimonious description of reality and are able to reproduce the observed data.

Other aspects, usually numerical parameters, are variable. A parameter inference step may be used to find the best-fitting value and adjust the model to its full potential. In cognitive science this step has occasionally been skipped in favor of hand-tuning a version of the model that produces the desired result, as was the case in the original publication of both the models presented in this thesis. One alternative approach to parameter inference is based on minimizing the error that the model produces based on some performance metrics chosen by the researcher. However, as the following section outlines, the most informative results may be obtained when the parameter inference step is firmly grounded in a statistically rigorous framework, e.g., by maximizing the probability of the observed data under the model, i.e., the likelihood of the model given the data (Myung, 2003).

The quality of the resulting fitted model should ideally be assessed on a test data set which is separate from the training data set used for parameter inference. Model quality is determined by how well the model predicts the target data. This may be

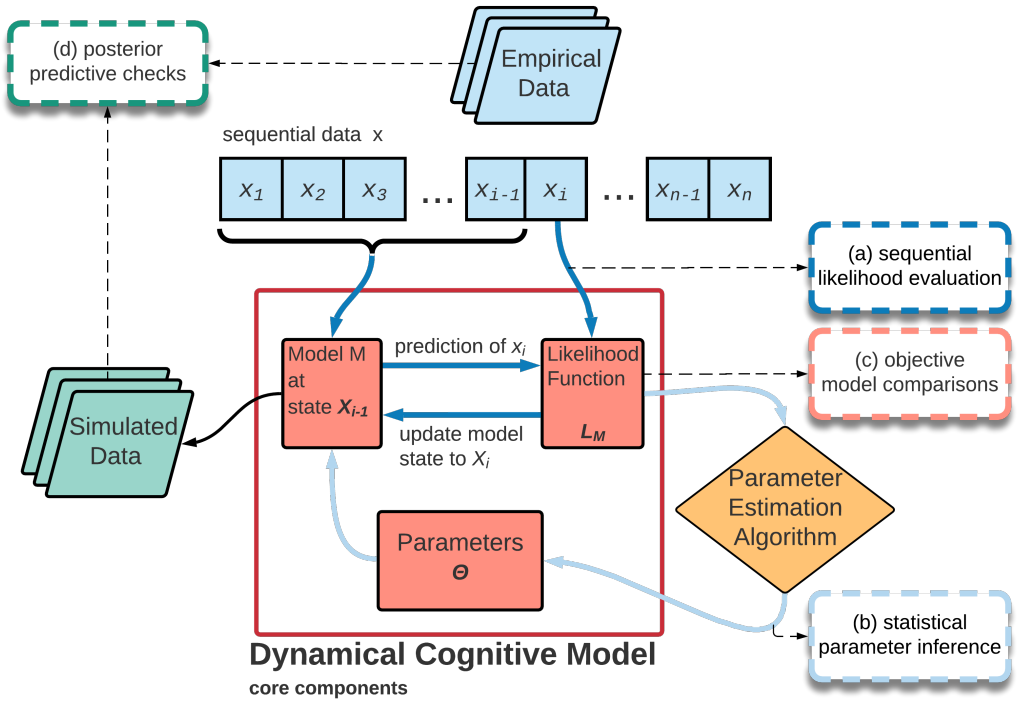

Figure 1.1 The likelihood-based cognitive modeling framework. This figure was originally published in a review paper by Engbert et al. (2022). The figure shows that a dynamical cognitive model takes sequential data as its input to update the internal model state. Thus each subsequent model state depends on the previous states. The model can compute the likelihood of the model given the data **(a)**, which may be used to conduct statistical parameter inference **(b)** and objective model comparisons **(c)**. Using the model in a generative way enables a comparison between experimental and simulated data, which we refer to as posterior predictive checks **(d)**.

measured generally by the model likelihood on the test set, but may also include more specific performance measures, relating to specific aspects of the data. The following section describes the modeling framework used throughout this thesis. The concrete case of likelihood-based dynamical cognitive models, which will be discussed in detail, is summarized in Figure 1.1. In a short review paper on the same subject we provide some additional examples (Engbert et al., 2022).

1.2.1 Likelihood-based modeling

Two popular ways of measuring how well a given model M with parameters θ describe some data X are (a) performance metrics and (b) statistical likelihood. Both are important for a comprehensive understanding of a model and are not interchangeable. In the context of parameter inference, the former is associated with minimizing a loss function, e.g., least-square estimation (LSE) and the latter with probabilistic methods such as Bayesian parameter inference and maximum likelihood estimation (MLE). As we will see, the likelihood-based approach is preferable for estimating parameters

whenever available because it is part of a rigorous statistical framework. In the context of model evaluation, it is fruitful to use both model likelihood and performance metrics.

Parameter inference based on performance metrics relies on a loss function, a metric that quantifies the difference between the experimental data and the model predictions. The precise nature of this loss function is highly dependent on the concrete use case, but requires the choice and formulation of specific metrics that represent the characteristic features of the target data. These metrics may be summary statistics or statistical tendencies. In order to infer optimal parameters, the model is used to simulate data, which is then compared to the experimental data using the chosen metrics. The more similar the characteristics of simulated and target data are, the better the model is considered to be. Using this approach to characterize the similarity between simulated and target data in a single value, it is possible to fit parameters, e.g., using a least-squares estimation.

Although this has been a popular choice in the field of psychology it has some significant drawbacks. First and foremost it relies heavily on the selected performance metrics. This choice is at least somewhat arbitrary and imposes strong constraints on what is considered important in the data (Kümmerer et al., 2015; Schütt et al., 2017). Moreover, it may involve implicit assumptions about the data, e.g., LSE assumes that the noise must be Gaussian. Fitting to specific performance metrics also reduces the informativeness of that metric in the later evaluation of the model: the model has been adapted to produce a specific behavior, potentially at the cost of other aspects that are invisible to the procedure. Furthermore, it lacks statistical rigour for testing hypotheses and does not enable estimating confidence intervals for the parameter inference. However, comparing model-simulated data with experimental data is an important component of the model analysis. Particularly in the case of procedural models, it is highly relevant to ascertain whether the intended behavior is being produced by the model and how the predictions differ.

An alternative approach uses the model likelihood. The likelihood of a model is defined as the probability $P(X|\theta)$ of observing some data X given a model with parameters θ and is most commonly written as $\mathcal{L}(\theta|X)$.[2] Not all models are designed to compute a likelihood. Probabilistic, likelihood-based models have the advantage of allowing statistically rigorous parameter inference using a Bayesian framework and allowing objective model comparisons.

1.2.2 Data assimilation

The models discussed in this thesis are both procedural, dynamical models that rely on time-ordered data. Methods to estimate parameters, compare and analyze dynamical models are well-established in other scientific disciplines (e.g., weather forecasting, see Asch et al., 2016) and are collectively referred to as Data assimilation (Reich & Cotter, 2015).

2 This formulation emphasizes that the likelihood is a function of the parameters, which is a more natural conceptualization for parameter inference.

The observations are given as a time-ordered sequence of n events $X_n = (x_1, \ldots, x_n)$. The probability of each event x_i in the sequence depends upon the sequence of preceding observations $X_{i-1} = (x_1, \ldots, x_{i-1})$, i.e., $P_{\mathrm{M}}(x_i \mid X_{i-1}, \theta)$. The likelihood of the model parameters $\theta \in \Theta$ given the data X_n is defined as the joint probability of a sequence of events X_n,

$$\mathcal{L}(\theta \mid X_n) = P_{\mathrm{M}}(x_1 \mid \theta) \prod_{i=2}^{n} P_{\mathrm{M}}(x_i \mid X_{i-1}, \theta) \ . \tag{1.1}$$

1.2.3 Parameter inference

Parameter inference is the process by which a researcher explores the range of values which its parameters may assume and finds which parameter values allow the model to best represent the data. Here, we focus on likelihood-based models, i.e., models that can compute the likelihood $\mathcal{L}(\theta|X_n)$ of some parameters θ given the data X_n.

In a Bayesian framework, we can compute the posterior distribution over the parameters using Bayes' Rule,

$$P(\theta \mid X_n) = \frac{\mathcal{L}(\theta \mid X_n) P(\theta)}{\int_\Theta \mathcal{L}(\theta^* \mid X_n) P(\theta^*) \mathrm{d}\theta^*} \ , \tag{1.2}$$

where θ is a vector of all model parameters and $P(\theta)$ is a vector of priors for the parameters. The posterior represents the full marginal distribution for each parameter. $P(\theta \mid X_n)$ is usually intractable, but there exist methods to approximate it. One key property of the posterior distribution is that the normalization term in the denominator, also known as the marginal likelihood or model evidence, does not influence the shape of the distribution, as it is not a function of θ. This makes it possible to use sampling algorithms to approximate the posterior distribution with just the likelihood and the priors, even when the marginal likelihood is difficult to compute.

A popular class of algorithms for sampling from a target distribution are Markov Chain Monte Carlo (MCMC) methods [3]. These algorithms are widely used for approximating the posterior distribution of the parameters of a statistical model given the data. There are many different MCMC algorithms available, each with its own strengths and limitations. The choice of algorithm often depends on the specific problem being solved. At a very basic level, MCMC samplers iteratively draw samples from the target distribution, using a Markov Chain to explore the parameter space. The samples are statistically correlated and the distribution of the samples converges to the target distribution as the number of samples increases. A notable method within MCMC algorithms is the Metropolis-Hastings algorithm (Hastings, 1970; Metropolis et al., 1953). Starting from an arbitrary point in the parameter space and each subsequent sample is generated by drawing from a proposal distribution (a probabil-

3 MCMC methods construct a Markov Chain, i.e., a sequence of random variables where the next state depends only on the current state, in order to sample from probability distributions.

ity distribution used to generate candidate values for the next state of the Markov Chain) and accepting or rejecting the point based on the target distribution. MCMC algorithms typically require a large number of samples to accurately approximate the target distribution. There are techniques for accelerating the convergence of the Markov Chain, such as using a more efficient proposal distribution or adapting the proposal distribution based on the samples generated so far. The details and comparisons of different performant algorithms for this process are outside the scope of this thesis. For the parameter inference of all the models developed in this thesis, I used an adaptive metropolis sampler.

MT-DREAM, i.e., multi-try differential evolution adaptive metropolis (Laloy & Vrugt, 2012), is a convenient choice for the parameter inference for the presented class of dynamical cognitive models (Schwetlick, Rothkegel, Trukenbrod, et al., 2020b; Seelig et al., 2020). First, it does not rely on any specific requirements, aside from the model likelihood which represents the target distribution. Second, the algorithm has been implemented and optimized as a Python package (Shockley et al., 2018), which has been shown to be fairly computationally performant. The MT-DREAM sampling algorithm implements three core principles: Adaptive Metropolis, multiple trial Metropolis, and differential evolution.

Adaptive Metropolis refers to the choice of proposal distribution, where instead of using a symmetrical Gaussian proposal distribution, as in classical Metropolis Hastings methods, the covariance of the proposal distribution is the covariance of the past (accepted) samples. The multiple-try aspect of MT-DREAM refers to the generation of multiple samples from the proposal distribution at once. This results in a lower rejection rate of samples and contributes to faster convergence. Lastly, MT-DREAM implements a method called differential evolution, which was first described by (Storn & Price, 1997). Differential evolution uses the the parallel sampling chains, modifying the acceptance/rejection of each sample dependent on the current state of the parallel chains (Braak, 2006).

A key advantage of Bayesian parameter inference is that access to the full marginal distribution for each parameter affords further valuable information. The marginal distributions for each parameter reflect the uncertainty of the estimate and can also give important information about characteristics of the model and the interdependence of parameters. Other methods of parameter inference simplify the parameter inference process by finding a point estimate for the parameters instead of approximating the full posterior distribution. Common point approximations are the Maximum A Posteriori (MAP) and the Maximum Likelihood (ML). Both are used to obtain a point estimate for the parameters that maximizes the likelihood.

In a Bayesian framework the MAP is defined as the maximum of the posterior distribution

$$\arg\max_{\theta \epsilon \Theta}(P(\theta \mid X_n)) = \arg\max_{\theta \epsilon \Theta}(P(X|\theta)P(\theta)) \; . \tag{1.3}$$

In a frequentist framework, the maximum likelihood is written as

$$(\hat{\theta}) = \arg\max_{\theta \epsilon \Theta} \mathcal{L}(\theta|\mathcal{X}) \ . \tag{1.4}$$

This is equivalent to the MAP with a uniform prior distribution $P(\theta)$. As described previously, in the case of complex models it is usually not feasible to analytically derive this maximum. Instead, it is necessary to numerically find the maximum by iteratively evaluating the model with varying parameters. As this is a computationally intensive but also highly relevant problem in many fields, there exist a plethora of algorithms that implement this process in various ways. One such algorithm is grid search, which works by dividing the parameter space into a finite number of grids and evaluating the model at each grid point. This approach quickly becomes computationally unfeasible, as the parameter space grows. Greedy optimization is another approach, which involves making a sequence of locally optimal decisions in the hope of finding a global optimum. The expectation maximization (EM) algorithm works by iteratively improving the estimates of the model parameters by using the expected value of the complete data log likelihood, which may include missing data. When the likelihood is continuously differentiable gradient-based optimization algorithms, such as gradient descent and conjugate gradient, offer another alternative. These algorithms make use of the gradient of the likelihood function and can often find the MLE more efficiently than other methods such as grid search or greedy optimization, especially when the objective function is smooth and has a single global optimum.

1.2.4 Model evaluation and posterior predictive checks

Once values or distributions for the free parameters of a model have been found, the model can be evaluated. The first measure to understand how well a given model represents the data is the model likelihood measured on a separate test data set. It is an important metric that allows fair comparisons to other models evaluated on the same data and gives an indication of how much of the variance in the data can be explained by the model, particularly when an appropriate comparison model is available.

However, the modeling framework also affords a more detailed analysis of the model predictions. These analyses are collectively referred to as posterior predictive checks. As stated by Kruschke (2014), "A posterior predictive check is an inspection of patterns in simulated data that are generated by typical posterior parameters values." For assessing the quality of a mechanistic model it is important not only that the model likelihood is high, but also that the estimated parameters which represent concrete interpretable quantities take on realistic values and the behavior is in good accordance with the empirically observed behavior.

The precise procedure of conducting posterior predictive checks depends highly on the research in question. Typically, the model is used in a generative way to simulate data. Then, a series of appropriate characteristics of the experimental data must be identified, which, ideally, should be present in the simulated data as well. While

ad-hoc metrics of specific aspects of the data should be viewed critically as a basis for parameter inference (Schütt et al., 2017), they are highly informative as a part of posterior predictive checks. When simulated data exhibits trends that were not explicitly built into the the model, it provides strong evidence for the proposed mechanisms in the model. Conversely, when certain statistical features of the empirical data are not present in the simulated data it can reveal potential for model extensions or modifications.

Posterior predictive checks, in the cases presented in this thesis, are often comparisons of statistical tendencies and distributions in the data, but may also include summary statistics. The goal, primarily, is to investigate whether the simulated data are typical in the space the experimental data. The results of posterior predictive checks serve to guide intuitions about which aspects of the model qualitatively succeed or fail. This intuition is valuable to generate ideas about how to address potential shortcomings in order to better capture the trends in the data.

1.3 Eye movement and vision

Vision is probably the most complex sense available to us. Every second $10^8 - 10^9$ bits of information enter the eyes (Kelly, 1962). Our typically seamless perception of the visual world belies the complexity of the signal analysis and filtering required to exploit visual information at this rate. This process begins with the architecture of the eye, including the receptor distribution; continues in early processing of information in the first layers of neurons which already relay summarized signals; and becomes ever more complex in later layers of the brain, detecting first contrasts and edges and eventually coherent visual representations. Visual attention, which usually coincides with the center of the visual field, is directed at relevant parts of the visual display in turn (Findlay & Gilchrist, 2003; Henderson & Hollingworth, 2003). Attention is guided by both bottom-up features of the scene, as well as top-down features of the task or prior knowledge.

The following section illuminates some of the fundamental principles of eye movement. As discussed above, the aim of this thesis is to build principled models of cognition where the internal workings of the model are associated with the proposed biological mechanism. The state of the experimental research is fundamental to the building of such models.

1.3.1 Fundamentals

Light entering the eye is first refracted by the transparent, curved cornea and aqueous humor. It travels through the lens which refracts and focuses the light, such that the focal point, after passing through the vitreous humor lands on the retina, i.e., a layer of photoreceptor cells. Muscles in the iris control the size of the pupil, i.e., how much light is let in, and a separate set of muscles deforms the lens to allow adjusting the focus to different viewing distances.

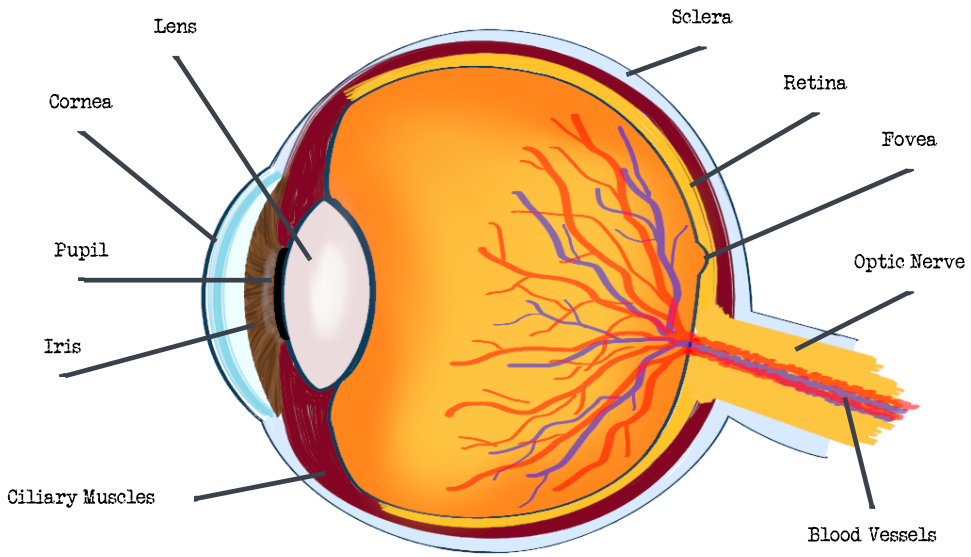

Figure 1.2 Labelled cross section of the eye. The eye focuses light onto the retina through the cornea and the lens. The retina contains photoreceptor cells that convert light into electrical signals that are sent to further visual processing areas in the brain via the optic nerve.

In the retina, photoreceptor cells turn the external stimulus into electrical signals in the nervous system. Photoreceptors can be divided into two categories: Rods are highly sensitive to light and contribute to vision primarily in low light conditions. Cones are more adapted to brighter conditions and are specialized to one of three bandwidth-ranges of light. This specialization enables color vision.

The first filtering mechanism of early visual processing is the distribution of photoreceptors in the retina. Receptor density concentrically decreases toward the periphery, limiting high visual acuity to the central fovea. It has been approximated that if acuity of the entire visual field were as high as in the fovea, the brain would need to be some hundreds of thousands times larger and weigh 10 tons to handle the information inflow (Findlay & Gilchrist, 2003). In the center of the visual field each individual photoreceptor connects to a single ganglion cell, which then transmits the information via the optic nerve into the visual processing centers of the brain. In the periphery of the visual field, a single ganglion cell may receive input from multiple photoreceptors, effectively summing their signal. Dedicated areas in the visual cortex are also proportional to this difference in sensitivity (Bear et al., 2007; Strasburger et al., 2011).

The foveated architecture of the retina requires active sampling of the environment by moving the eyes in order to build a complete visual representation. The eyes explore the visual world in a sequence of ballistic jumps, saccades, and periods of relative stability known as fixations. Other types of eye movement include smooth pursuit

which occurs in response to movement in the environment and vestibulo-ocular eye movements which correct reflexively for head movement. Movements in the human eye are controlled by six extraocular muscles, four in the cardinal directions and two oblique (Spencer & Porter, 2006). The bulk of eye movement research focuses on fixations and saccades.

1.3.2 Scan paths

Sequences of fixations and saccades are known as scan paths. The high-resolution fovea is directed at different areas of the visual world in turn. Studying the selection process of fixation locations can reveal much about the underlying visual processing. The concept of visual attention is of high relevance in this context. During typical viewing the locus of visual attention and the eye position coincide, in what is known as overt attention (Findlay & Gilchrist, 2003). The rarer case is when visual attention and fixation position are not aligned, as covert attention. This is typically only the case when intentionally not moving the eyes, but is also relevant briefly before and after a saccade.

A passive interpretation of vision asserts that the eyes sample the visual world in order to build an internal model, which may serve as a basis for action planning (Aloimonos & Rosenfeld, 1991; Marr, 1982). On the contrary the current prevailing theory is that a rich internal model is superfluous. The constant availability of the real world as a rich reference allows an active vision interpretation, where the eyes simply move to the elements of interest as they become relevant (Findlay & Gilchrist, 2003; O'Regan, 1992). Although active vision, and eye movement, is widely accepted to play a formative role in visual perception, much of vision research focuses on static paradigms. Active vision emphasizes the interplay between perception and action and is therefore a highly compatible with the dynamic modeling framework presented in this thesis.

1.3.3 Saccades

Saccades occur at a rate of 2-5 saccades per second, although this number is highly dependent on the task and on inter-individual variation. The same is true of the typical distance a saccade covers, i.e., the saccade amplitude: during reading a typical saccade covers 1-2° visual angle, whereas during scene perception this distance is around 4°, as reported by (Rayner, 1998). The typical saccade amplitude varies from one experimental paradigm to the next, with influences including the size of the presented stimulus (von Wartburg et al., 2007), the task (Backhaus et al., 2020), and the image feature distribution in the image.

The actual saccade movement is fast and precise, minimizing the reduction of visual input during the movement and allowing a seamless perceptual experience and quick reactions to visual stimuli. In order to achieve this, the saccade motor program is stereotyped and (semi-) ballistic in nature (Harris & Wolpert, 2006). The preparation of a saccade to an experimentally defined location takes about 150 ms (Rayner

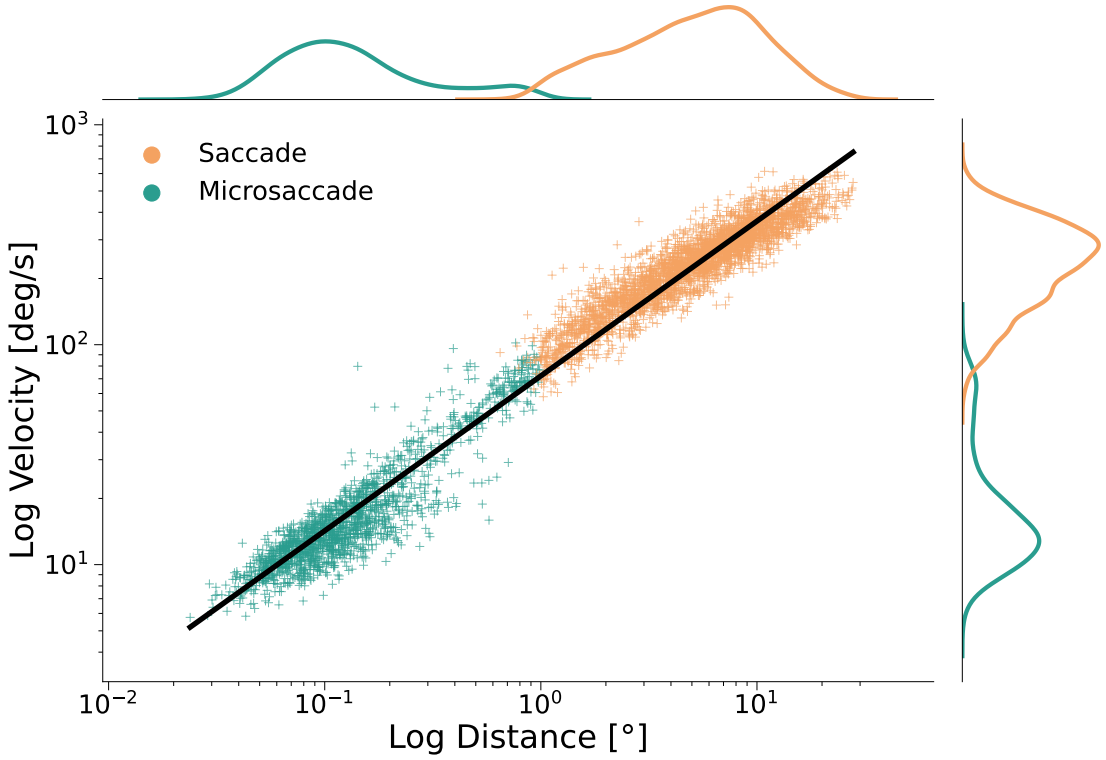

Figure 1.3 Main Sequence of saccades. There is a characteristic linear relationship between the logarithmic saccade amplitude and the logarithmic peak velocity of the saccade. Events colored in orange represent regular saccades while events colored in green are considered microsaccades.

et al., 2009). During the first part of the preparation (first 100 ms) saccades may be modified or even canceled; then, a point of no return is reached, where the saccade is inflexible to further changes (Becker & Jürgens, 1979). As such, saccades have a stereotypical velocity and movement profile. One characteristic is that the logarithm of the distance travelled is linearly related to the logarithm of the maximum movement speed. This relationship, known as the Main Sequence, is shown in Figure 1.3 (Collewijn et al., 1988). Saccades that cover less than 1° of visual angle are typically considered microsaccades (see Section 1.3.5).

The classical interpretation of saccades posits that during the movement no visual information can be registered, a phenomenon called saccadic suppression (Matin, 1974). More recent research suggests that the reduction in visual processing during saccades is more the consequence of an absence of appropriate information than a deliberate suppression. When the stimulus is stabilized on the retina, or happens to be appropriate for the velocity during the saccade, visual perception is possible during saccades (Castet & Masson, 2000). For example, certain moving sine gratings, may appear as monotonous surfaces during fixation but can be perceived during saccades (Castet & Masson, 2000; Mathôt et al., 2015).

Despite the fact that the eyes move around sampling potentially disjointed areas of the environment, the final perception of the visual world is smooth and seamless.

In order to maintain this illusion of constancy and coherence, saccade planning must include assembly of inputs relative to previously viewed locations. While naturally during fixations visual attention and fixation location are aligned (Findlay & Gilchrist, 2003), around the time of a saccade studies have measured increased processing in other locations. For example just before a saccade, processing at the upcoming target location is enhanced (Deubel & Schneider, 1996; Irwin & Gordon, 1998; Kowler & Blaser, 1995). In what may be called predictive attentional targeting, covert attention precedes the eyes to their target (Posner, 1980) as early as 150 ms before saccade onset (Rolfs et al., 2011). Furthermore for 100-200 ms after the execution of a saccade, processing benefits may be measured in the retinotopic position that the target was in, before saccade execution (Golomb et al., 2008). This retinotopic attentional trace indicates that visual attention moves retinotopically with the saccade (Marino & Mazer, 2016). Thus, the visual system anticipates and predicts the changes in input information and reallocates resources to anticipate changes in input and to allow the visual experience to be seamless.

1.3.4 Fixations

The process of choosing each fixation location in the sequence relies on processing the available information and identifying the most interesting or promising target. Research as early as 1935 (Buswell, 1935; Yarbus, 1967) showed that some areas are more likely to be fixated than others. This choice depends on a variety of attentional biases.

Bottom-up, or stimulus influences on the choice of fixation locations include image features such as edges and luminosity (Itti et al., 1998; Mannan et al., 1997; Tatler et al., 2005). They may reflect evolutionary preferences for certain statistical aspects of an image (Itti & Koch, 2001) and are well-understood to form the basis for visual processing. Luminance and color differences, for example, are responded to by ganglion cells in very early layers of visual processing (Kolb et al., 2001; Polyak, 1941; Purves et al., 1997). In the primary visual cortex cells may be found that respond to specific orientations and spatial frequencies (Blakemore & Campbell, 1969; De Valois et al., 1982; Schütt et al., 2017). The degree of complexity increases in later processing stages of the visual system, from basic features, to the detection of edges and contrasts, to meaningful objects and relationships between objects. Positive correlations have been found that relate fixation positions and image statistics (e.g., Hallett, 1978; Privitera & Stark, 2000; Reinagel & Zador, 1999; Tatler et al., 2005; Theeuwes et al., 1998).

Top-down influences like task requirements, experience and cognition also strongly influence the choice of fixation position (Backhaus et al., 2020; Castelhano et al., 2009; DeAngelus & Pelz, 2009; Hayhoe et al., 2003; Yarbus, 1967). One memorable study compared expected and unexpected objects in a scene, i.e., a picture of a farm containing a tractor or an octopus. The unexpected object was fixated more frequently and for longer, despite controlling for feature saliency (Loftus & Mackworth, 1978).

Some other attentional biases are independent of the image and of the task (Foulsham & Kingstone, 2010; Foulsham et al., 2008; Schütt et al., 2017; T. J. Smith &

Henderson, 2009; Tatler, 2007). A prominent example is the central fixation bias: the center of an image receives more and longer fixations, particularly in the first few fixations (Rothkegel et al., 2017). Further biases include a preference for the cardinal directions and a variety of sequence effects, like a preference to continue along the saccade vector rather than change direction (Rothkegel et al., 2018; Rothkegel et al., 2016; T. J. Smith & Henderson, 2009; Tatler & Vincent, 2009).

Fixation duration differs between individuals (Henderson, 2003) and with the given task and stimulus. The average fixation duration has a mean length of 300 ms and their distribution resembles a right-skewed Gamma distribution (Henderson, 2011; Henderson & Hollingworth, 1998). Fixation durations are typically interpreted as an indication of ongoing information processing (Rayner et al., 2009). A discussion in the literature surrounds whether they are controlled indirectly or directly (Henderson & Smith, 2009; Trukenbrod & Engbert, 2014). Direct control refers to the influence of current visual input on the length of a fixation (Rayner, 1995), while indirect control refers to factors such as an autonomous timer that independently triggers a saccade after a random time interval (Engbert & Kliegl, 2001; Nuthmann & Henderson, 2010). Finally, in mixed control, a combination of direct and indirect factors determine fixation durations (Henderson & Pierce, 2008).

1.3.5 Fixational eye movement

Even during the relatively stable fixations the eyes are not still. A family of microscopic eye movements perturb the fixation position: Tremor, Drift, and Microsaccades (see Martinez-Conde et al., 2004). The current state of understanding posits that these movements prevent image fading due to neural adaptation (Kowler, 2011; Martinez-Conde et al., 2004; Martinez-Conde et al., 2006), i.e., a reduction in the firing rate of neurons due to continued exposure to a stimulus. Additionally, visual acuity, particularly for fine spatial vision, is also related to fixational eye movement (Rucci et al., 2007).

Tremor is the smallest amplitude movement with amplitudes of only 0.1 to 0.5 minarc and frequencies of around 90 Hz (e.g., Adler & Fliegelman, 1934; Higgins & Stultz, 1953; Ratliff & Riggs, 1950). Even using state of the art video-based technology, it is difficult to accurately measure tremor. It is likely to originate in low-level but central areas connected to the extra-ocular muscle motor units and is correlated between the eyes (Spauschus et al., 1999).

The second category of fixational eye movement is drift. The eyes meander in a seemingly random way around the fixation position. Their trajectory may be described as a self-avoiding random walk (Engbert et al., 2005). Drift movement, like tremor movement, is difficult to measure accurately and this may be the reason why early studies found no correlation between the eyes (Krauskopf et al., 1960; Yarbus, 1967). Following studies (Spauschus et al., 1999) did report binocular coherence and (Thiel et al., 2006) showed significant phase synchronization of the movement. Note that the movement is not necessarily identical in both eyes, but is coordinated with regard to velocity and movement direction. Drift movements are also associated with enhanced

visual acuity for fine spatial vision, providing a more effective way to sample high-frequency image content (Rucci et al., 2007).

Lastly fixational eye movements include microsaccades. Microsaccades occur at a rate of 1-2 per second and are similar to regular saccades in all but amplitude: they are significantly different from drift movements regarding their velocity, typically occur in both eyes (Engbert & Kliegl, 2003) and their movement profile is well-described by the Main Sequence (Zuber et al., 1965), as shown in Figure 1.3). A typical definition considers all saccades of less than 1° of visual angle a microsaccade. Microsaccades occur spontaneously and mostly involuntarily, although the rate and direction may be altered by various circumstances. For example evidence shows that they preferentially occur in the direction of covert attention (Engbert & Kliegl, 2003). In an oddball paradigm, the rarer stimulus (or oddball) has been found to inhibit microsaccades (Valsecchi et al., 2006). A more functional interpretation is that microsaccades are corrective for drift (Martinez-Conde et al., 2004) or blinks (Costela et al., 2014).

Like drift, microsaccades are related to visibility. While perceived fading does not causally trigger microsaccades (Poletti & Rucci, 2016), periods of perceived fading of a peripheral stimulus is associated with a decrease in probability, rate, and magnitude of microsaccades while transitions toward visibility are associated with an increase (Martinez-Conde et al., 2006). Drift by itself already effectively prevents perceptual fading (Collewijn & Kowler, 2008), however while prevention of fading may be achieved by just drift, microsaccades are much more effective at restoring vision after fading (McCamy et al., 2014; McCamy et al., 2012). An additional layer of complexity is added by the fact that peripheral receptive fields are considerably larger than foveal receptive fields; while drift alone may effectively prevent foveal fading, microsaccades may be necessary to prevent peripheral fading. Microsaccades have also been found to contribute to the disambiguation of latency and brightness, as well as helping to synchronize and modulate the summation of neurons with neighboring receptive fields (Martinez-Conde et al., 2004). Thus, the role of microsaccades for visual perception seems to be related to low level visual processing, fixational control, and attentional processes.

Broadly we can identify two major research questions to study visual processing. First, on a microscopic level: how do features get extracted and combined from photoreceptor activity and how do eye movements contribute? Second, taking such feature extraction as a basis: which components contribute to the active sampling that is eye movement to decide when and where to look? The following chapter will give an overview about the modeling literature in these fields.

1.4 Modeling macroscopic gaze behavior on natural scenes

Due to its complexity and high relevance for interacting with the world, the field of vision science comprises a vast number of models that describe parts of the visual processing stream, at various levels and with various goals. As eye movements drastically control visual input, understanding the processes that guide eye movement is

a promising approach to understanding visual perception. Models range from highly specialized domains like reading (Engbert et al., 2005; Reichle et al., 1998), driving (Chapman et al., 2002; Land & Lee, 1994; Underwood et al., 2003; Underwood et al., 2007), viewing faces (Haxby et al., 2002; Peterson & Eckstein, 2012, 2013), or playing chess (Reingold & Sheridan, 2011) to more general contexts (Adeli et al., 2016; Engbert, Trukenbrod, et al., 2015; Itti & Koch, 2000; Treisman & Gelade, 1980; Trukenbrod & Engbert, 2014; Tsotsos et al., 1995). Eye movements are also regarded as an observable (overt) consequence of covert visual attention. Therefore, models of attention and models of eye movement are often presented jointly.

In the first part of this thesis I focus on models of eye movement in response to photographs of natural scenes in a laboratory environment. The reason for this focus is twofold. First, there exists a large literature researching scene viewing behavior in this context. This also means that expertise and high quality experimental data are available. Second, viewing static scenes restricts the types of eye movement that can be observed to fixations and saccades. When eye movements are recorded in more complex circumstances, e.g., when the participant performs other actions (Land et al., 1999) or in response to moving video stimuli, the range of observable eye movement types expands to include vergence and accomodation eye movements, as well as motion tracking, and defocus (Meese et al., 2006). These types of movement are experimentally less thoroughly studied and add a large amount of complexity to the model and to the computations.

1.4.1 Modeling fixation locations

Eye movements on photographs are characteristic in the sense that some areas are more likely to be looked at than others (Buswell, 1935; Yarbus, 1967). By recording people's eye movements and aggregating them over time, it is possible to compute a fixation density map (see Figure 1.4), which shows the preferred regions. A purely qualitative examination already reveals the unsurprising fact that people, in the absence of a task, look at regions in the image that contain interesting content such as objects or faces. These regions of interesting content are often referred as *salient* and the spatial distribution of salient regions as a *saliency map*. Initially the term *visual saliency* was used to refer mostly to low-level image features like edges (Itti et al., 1998). Later, the concept was expanded to include higher-level features such as objects or faces (Kümmerer et al., 2018).

The last years have seen a range of models that aim to compute, on the basis of image information, where these preferred regions are likely to be. The first models of this kind were based on research that correlated fixated locations with spatial features such as higher spatial contrast (Mannan et al., 1996, 1997) and other similar properties (Krieger et al., 2000; Parkhurst & Niebur, 2003; Reinagel & Zador, 1999). Simultaneously, in visual search research it was remarked that certain features would make a target visually "pop out" from among a range of distractors. In Feature Integration Theory, Treisman and Gelade (1980) hypothesized that the visual attention system operates in two steps. First, maps of basic features are computed in parallel

over the whole display. Then, attention moves and serially focuses on specific areas to understand more complex conjunctions of features. This theory was implemented computationally (Itti et al., 1998) in a model that extracts image features (color, intensity and edges, enter-surround differences). While this model was not intended to represent eye movements (it posited covert shifts in attention), it turned out to predict above-average model-saliency for fixated locations (Parkhurst et al., 2002). There is ever-growing range of models that predict saliency and fixation locations (Harel et al., 2006; Kienzle et al., 2009). The MIT/Tuebingen Saliency Benchmark (Judd et al., 2012; Kümmerer et al., 2018) keeps a list of these models and their performance. Overall the results show that purely feature-based models struggle to make accurate predictions of fixation locations (Bylinskii et al., 2016) and that adding higher level features is beneficial (Einhäuser, Spain, et al., 2008; Judd et al., 2009). Most notably the DeepGaze II model (Kümmerer et al., 2017), is a deep neural network that predicts fixation locations with unrivalled accuracy, using an image classification network as a basis which includes high- and low level features. For a more extensive overview of saliency models see, e.g., Borji and Itti (2013), Borji et al. (2013), and Kümmerer et al. (2018).

The level at which a model extracts information from an image to determine its interestingness, has recently been subject to much debate. High level features like faces and objects tend to heavily overlap with low level image features such as edges and contrasts, making it difficult to disentangle satisfactorily. At one end of the spectrum are purely feature-based models of classical saliency (e.g., Harel et al., 2006; Itti et al., 1998). These classical models have a strong link to early visual processing (Itti & Koch, 2000; Li, 2002). On the other hand theories of top-down guidance of behavior claim that the major influences are task while viewing the scene (Einhäuser, Rutishauser, et al., 2008; Henderson et al., 2007; Matthis et al., 2018; Pelz & Canosa, 2001), expectations (Cornelissen & Võ, 2017; Henderson et al., 1999; Loftus & Mackworth, 1978) or gist extraction (Torralba et al., 2006). Top-down influences are not as easily formalized and models that include these ideas directly are rare (a simplified but noteworthy exception is presented by Torralba et al., 2006). However, generally, the inclusion of high level features improves predictions (Borji et al., 2013; Judd et al., 2009; Kümmerer et al., 2017). In the DeepGaze II model (Kümmerer et al., 2017), a deep neural network, the feature extraction network is based on a training that favors high level features and object. This approach has proven to be highly successful in predicting fixation locations.

Another approach has proposed to omit low level influences all together, constructing fixation predictions from image patches rated for their meaningfulness (Henderson & Hayes, 2017, 2018). The authors claim that this metric predicts fixations better than state of the art saliency models, a result that has been heavily challenged (Pedziwiatr et al., 2021a, 2021b). Effectively, the question that is proving hard to answer is whether people fixate areas including interesting things, which happen to be highly correlated with image features, *or* whether eye movements tend to fall on areas with many image features, because this is a good strategy for finding areas which will turn out to have meaning. This chicken-or-egg problem remains as an open research ques-

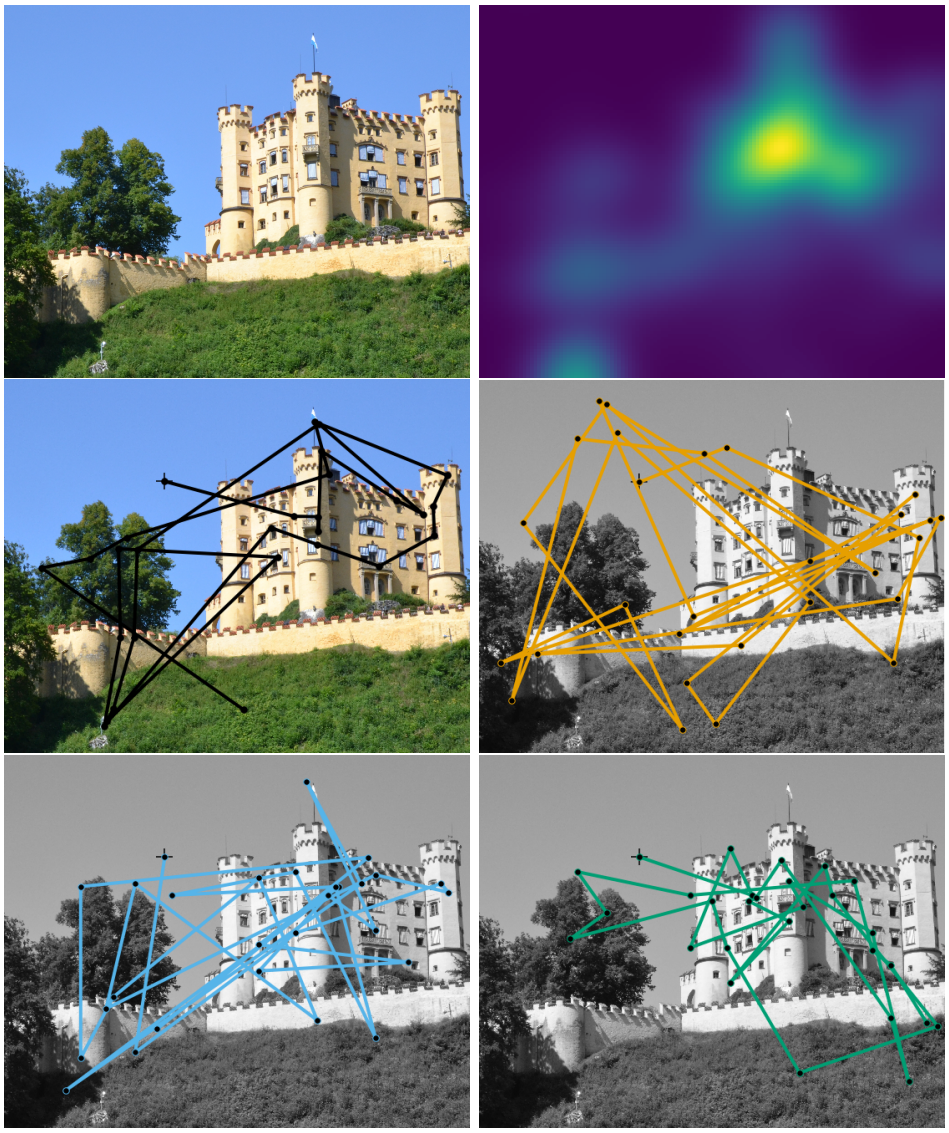

Figure 1.4 Real and simulated eye movement trajectories. Moving from the top left to the bottom right: the plain image; the experimentally computed fixation density map of the image; a real eye movement trajectory recorded from a person (black); a random trajectory, based on a uniform probability (yellow); A trajectory based on a static saliency model (blue); A simple dynamic model combining static saliency with a local focus (green).

tion and may require different research approaches to resolve. It seems unlikely that a satisfying consensus can be found by evaluating fixation locations alone.

1.4.2 Modeling dynamical and sequence effects

A fact that is often ignored in the research of fixation locations and image saliency, is that vision is not a static process. As the eyes move over a scene, the relevance of the existing static image features changes dynamically. High fidelity information is available only at close range to the current fixation position, so the knowledge about the image at any point in time is incomplete. The choice of fixation location therefore depends on a range of biological limitations and attentional control mechanisms. While a static saliency model may be performant for predicting locations, the actual sequence of eye movements is not well-represented. Figure 1.4 shows fixation sequences generated by a random model, a state of the art saliency model and an empirical scan path. The qualitative differences are immediately apparent. However, this can hardly be considered a shortcoming of the saliency model, which performs its intended purpose of identifying salient regions very well. It does, however, limit its capability for describing or explaining human behavior. In order to adequately describe sequences of eye movements more constraints are necessary.

Each eye movement involves a decision about when and where to move, presumably to maximize the relevant information to be gained. Early research on the neurophysiology of eye movements indicates that the decisions about when and where to move are partially independent systems (Carpenter, 2000; Findlay & Walker, 1999b). However, increasing evidence shows that, behaviorally, fixation duration and location are related. For example, regions that are fixated more frequently also tend to be fixated for longer periods of time (Einhäuser & Nuthmann, 2016) and fixations that precede a saccade to a highly salient locations tend to be shorter (Tatler et al., 2017). Furthermore, there is a tendency to maintain the direction of saccadic movement, especially after short fixations (T. J. Smith & Henderson, 2009). More time-related effects exist at the level of the sequence: there is evidence for an overall coarse-to-fine strategy (Over et al., 2007; Trukenbrod et al., 2019), a central fixation bias at the beginning of a sequence (Rothkegel et al., 2017; Tatler, 2007), an inhibition of return followed by a facilitation of return (T. J. Smith & Henderson, 2009). Thus, the dimension of time, i.e., both the duration and order of fixations, is crucial to understanding visual processing.

Taking these results seriously has led to the development of models of fixation sequences. Even a simple model of local saliency, i.e., that constrains each saccade length by weighting a saliency map relative to the current location significantly improves predictive power (Parkhurst et al., 2002). Implementing more of the systematic attentional processes and biases described above into sequence models has proven to be a promising step toward generating more realistic scan paths. One example of this shows that the concept of saccadic flow, i.e., a combination of systematic biases, can be linked to the saliency at the following location (Clarke et al., 2017). Models of scan paths typically focus on specific aspects of the eye movement and optimize for those

specific metrics, such as saccade distributions (e.g., Boccignone and Ferraro, 2004) or fixation durations (e.g., Nuthmann & Henderson, 2010). When modeling a specific task, such as visual search, metrics such as search efficiency become additionally available as model performance metrics (Zhou & Yu, 2021).

However, as discussed in Section 1.2.1, the full potential of the modeling approach is best explored when applying rigorous statistical inference. An early example was presented by Brodersen et al. (2008), who used a Bayesian modeling framework to construct a rise-to-threshold model of eye movement and learning in response to artificial stimuli. A later, methodologically important advancement suggested a Hamiltonian Markov Chain Monte Carlo approach to modeling scan paths using feature-level saliency, semantic content, and spatial position, and is notable as one of the first models of scan paths to use an explicit likelihood function (Liu et al., 2013). Another example uses Hidden Markov Models and variational inference to represent individual differences in the scan paths of different subjects (Coutrot et al., 2017). From the large range of models of fixation sequences developed in recent years, I will highlight five models in the following paragraphs, that are particularly relevant in the context of this thesis.

A notable example of a model that predicts scan path statistics was published by Le Meur and Liu (2015). The model takes a static saliency map (computed using a model by Harel et al., 2006) as a basis and combines it with distributional assumptions about saccade amplitude and directions which are fitted to the experimental data. Fixation locations are generated from this target selection map, under inclusion of an inhibition of return mechanism which is active over 5 fixations. This model reproduces saccade amplitude and angle distributions and performs highly in a saliency model comparison. An extended version of the model also fitted differences between semantic visual categories (Le Meur & Coutrot, 2016). A caveat of this model is that the model inference, i.e., the fitting of distributional assumptions, is not standardized, and that the underlying assumptions are more data-driven than theory-driven.

A different approach is implemented in the LATEST model (Tatler et al., 2017), which combines spatial and temporal aspects of eye movement. The proposition of the model is that fixation durations are related to the extraction of visual information, triggering a movement when the evidence for new location outweighs the present location. Thus, each event duration is modeled in a process-oriented way, based on a rise-to-threshold process (Reddi & Carpenter, 2000). The accumulation rate is tied directly to data-driven distributional assumptions. The LATEST model demonstrates that information accumulation processes fitted exclusively to durations can make valid predictions for fixation locations. The fitting procedure for this model is based on linear mixed effects modeling.

Another model that combines spatial and temporal eye movement control is the WALD-EM model (Kucharsky et al., 2021). It combines information accumulation using a drift-diffusion component (P. L. Smith & Ratcliff, 2004) and a spatial likelihood based on systematic attentional tendencies in the data. This model is particularly notable for the use of a combined spatiotemporal likelihood function and statistically rigorous parameter fitting. WALD-EM successfully reproduces a number of scan path

statistics and individual differences.

A model that explicitly represents the dynamic spread of attention is presented by Zanca et al. (2020). Their model uses differential equations inspired by the laws of mechanical physics. The four free parameters of the model are fitted using the normalized saliency at the fixated locations as a performance metric. This model is of interest because it predicts both saliency and scanpaths, and therefore does not rely on precomputed saliency maps. It has also been applied to predict eye movements on videos in addition to static photographs, demonstrating a noteworthy flexibility of application.

In this thesis I implemented and extended the SceneWalk model, a dynamical model of fixation sequences, which will be discussed in great detail in Chapter 2 and 3. SceneWalk sets itself apart by (1) being based on first order principles that are known about the physiology of the visual system, such as foveation and theories about the deployment of attention around the time of a saccade, and (2) by being a dynamical model, in the sense that the predictions evolve continuously over time instead of discretely. The premise is to explore whether these mechanistic constraints produce the statistical tendencies found in human eye movement data.

1.5 Modeling microscopic fixational eye movement

Fixational eye movements form the ubiquitous basis for all of visual perception. In contrast to macroscopic eye movements, fixational eye movements are to a much lesser extent under conscious control and its function is less clearly understood. A critical discussion concerns whether fixational movement has a functional role for visual perception. The alternative view posits that it is a necessary nuisance caused by imperfect muscular control or by the need to prevent visual fading, which needs to be corrected for further down the processing stream without additional benefits.

1.5.1 Models of fixational drift

During fixations the eyes drift smoothly and randomly around the fixation position. This fixational drift resembles Brownian motion over short periods of time (Burak et al., 2010; Engbert et al., 2011; Pitkow et al., 2007). This movement is characterized by increasing variance over time, which can be shown by calculating the the mean square displacement (MSD) at different time intervals. While Brownian motion has a linear increase in MSD, fixational drift exhibits a tendency towards persistence (Metzler & Klafter, 2000). Over longer intervals, the drift becomes antipersistent, meaning that it does not deviate further from the fixated location (Engbert & Kliegl, 2004). Modeling of fixational drift has proved to be an effective means of investigating its origin and role in visual processing.

The earliest model of fixational eye movement was proposed by Eizenman et al. (1985). The authors suggested a Poisson process representing random neural firing of a motor unit generates ocular tremor. Fixational drift is explained as a secondary, cyclo-

stationary process, which emerges from the summation of signals from different motor units firing at distinct frequencies. This model proposes that the motion frequencies found in tremor and drift are consistent with their origin in the individual motor units. However, more recent work on fixational eye movements places their origin higher up the chain of command, in the oculomotor integrator. The authors Ben-Shushan et al. (2022) developed a model based on the physiology of the primate visual system, using experimentally obtained estimates for parameters such as number of neurons, their tuning curves, and their spiking variability. The results of the model and the presented experimental findings are consistent with an upstream source of fixational drift upstream jointly informing the oculomotor neurons.

Experimentally, it has been found that fixational eye movements are not only associated with the prevention of visual fading (Kowler, 2011) but also play a role in enhancing visual acuity (Rucci et al., 2007). As fixational eye movements perturb the visual input, at a minimum mathematical models of visual perception need to account for fixational drift (Burak et al., 2010; Pitkow et al., 2007). More recent research shows that combining the edge detection properties of retinal neurons and fixational eye movements can lead to more robust edge detection than the neuronal properties alone (A. G. Anderson et al., 2020; Schmittwilken & Maertens, 2022). In the presence of fixational eye movements the features of the stimulus move over the individual receptors. This is in good agreement with the finding that neurons respond best to changes in signal and, specifically, that the visual system responds best to luminance transients. Moreover, the temporal component transports valuable information regarding orientation. A model with no orientation components, nonetheless accurately performs the edge detection task in the inclusion of a movement component (Schmittwilken & Maertens, 2022). These results indicate that fixational drift is not a nuisance factor to account for but a functional component in high acuity vision.

Models relating to contributions of drift to visual acuity typically assume a simple random drift. However, as presented above, experimentally measured fixational drift contains some systematic statistical properties such as the transition from persistent to anti-persistent movement and the distribution of angles. These properties can be used to predict more realistic drift trajectories and may point towards the mechanisms causing fixational drift. A recent paper by A. G. Anderson et al. (2020) proposes a Bayesian modeling approach that integrates the inference of the movement and the stimulus. The authors present experimental eye movement data, collected in response to a set of pattern stimuli. The model aims to simultaneously predict the movement and the stimulus pattern. It assumes that retinal cells are arranged in a grid and the patterns are projected onto these cells. The cells fire at some rate representing the way the pattern falls on the retina. In order to determine the stimulus from the spike rate, the authors propose a Bayesian formulation that computes the probability of the stimulus given the observed pattern of spikes. Using alternating steps, the stimulus and the movement estimates are fixed, inferring first one, then the other. The underlying idea is that the pattern cannot be estimated without accounting for movement and the movement can not be understood without a representation of the pattern. In the model there is no efference copy of the movement. The model results

are consistent with the idea that drift motions benefit high acuity vision, mainly by averaging over the inhomogeneities in the retinal receptors and receptor density.

A generative model of fixational eye movement was proposed by Engbert et al. (2011). The SAW model is a self-avoiding random walk inside a constraining potential. The self avoidance is implemented by activating visited locations on a grid and preferentially moving to less active locations. These activations decay according to a differential equation, which represents the memory of the process. The SAW model is able to reproduce the persistent and anti-persistent properties of fixational drift. The addition of neurophysiological delays was shown to also reproduce the characteristic oscillations found in the displacement autocorrelation (Herrmann et al., 2017). This model does not depend on any stimulus properties and focuses solely on accurately modeling the statistical properties of the movement. In Chapter 4 I implemented the SAW model and applied a likelihood-based parameter inference for each individual subject in a data set in order to investigate interindividual differences in behavior.

Another model of fixational drift that implements a self-avoiding random walk was suggested by Roberts et al. (2013). This model chooses each step direction from a continuous distribution that is weighted by the density of recent movement directions. The authors motivate the self-avoiding mechanism as a way avoid neural adaptation. Stronger neural responses to the novel transient stimuli are caused by the movement. The model shows the expected persistent behavior at short time scales, but has no mechanism that would account for the change to anti-persistent behavior.

1.5.2 Microsaccade models

Microsaccades, like drift, have been suggested to contribute to the prevention of fading. The relative contribution of the types of movement is debated, but it has been suggested that drift continually prevents fading and microsaccades effectively reverse it (McCamy et al., 2014). Microsaccades have also been found to contribute to visual acuity by precisely repositioning the eye (Ko et al., 2010) and contributing to edge detection, as suggested by mathematical model of microsaccadic displacement (Donner & Hemilä, 2007).

The exact mechanism responsible for triggering microsaccades is still not fully understood. Several mathematical models have been proposed over the years to explain the generation of microsaccades, but a comprehensive understanding of this process remains elusive. Microsaccade occurrence, which is highly individually variable, can be understood as a Poisson Process with individually different rates (Engbert & Mergenthaler, 2006). A common idea, which has also found experimental support (Otero-Millan et al., 2013), is that microsaccades exist on the same continuum as regular saccades. The same underlying circuits in the brain are responsible for executing saccades of all sizes (Martinez-Conde et al., 2013). However, large saccades and microsaccades do show distinct descriptive properties and functional characteristics (Mergenthaler & Engbert, 2010).

In a theoretical model proposed by Rolfs et al. (2008) for the generation of saccades and microsaccades, a motor map in the Superior Colliculus represents activa-

tion coding for both fixation and saccades. The central site represents fixation and activation in the peripheral regions triggers saccades with amplitudes proportional to the distance. The activity at the central, fixation-related site in the map predicts the frequency, amplitude, and direction of microsaccades. An alternative model with a neurophysiological focus (Otero-Millan et al., 2011) suggests that a circuit composed of omnipause and long-lead burst neurons driven by activity and in the superior colliculus triggers both microsaccades and saccades. The model operates based on the reciprocal inhibition between the omnipause and long-lead burst neurons, triggering an eye movement whenever the long-lead burst neurons overcome the inhibition from the omnipause neurons. Thus, this work suggests a common triggering mechanism for regular saccades and microsaccades.

Other evidence suggests that that microsaccades are related fixational drift. Potential triggering mechanisms that have been proposed include low retinal image slip, i.e., a reduction in fixational drift (Engbert & Mergenthaler, 2006). With respect to the SAW model, Engbert et al. (2011) also propose a mechanism for for generating microsaccades based on the activation in the SAW model. In Chapter 4, I investigate this idea further by investigating the relationship of internal model activation states and microsaccades.

The following chapters present two different models of eye movement behavior. Chapter 2 and 3 report on the implementation and extension of the SceneWalk model of scan paths. The addition of attentional mechanisms around the time of a saccade is shown to improve scan path predictions. Additionally, we discuss insights from leveraging the potential of the Bayesian parameter inference to fit individual subjects and tasks. Chapter 4 presents a likelihood-based implementation of the SAW model of fixational eye movement. We discuss results from individual subjects and investigate the connection between fixational drift and microsaccades. These models of eye movement behavior serve as examples for the application of the described modeling framework to the field of cognitive modeling.

2 Modeling perisaccadic attention

Modeling the effects of perisaccadic attention on gaze statistics during scene viewing

Lisa Schwetlick, Lars Oliver Martin Rothkegel, Hans Arne Trukenbrod, Ralf Engbert
University of Potsdam

Abstract

How we perceive a visual scene depends critically on the selection of gaze positions. For this selection process, visual attention is known to play a key role in two ways. First, image-features attract visual attention, a fact that is captured well by time-independent fixation models. Second, millisecond-level attentional dynamics around the time of saccade drives our gaze from one position to the next. These two related research areas on attention are typically perceived as separate, both theoretically and experimentally. Here we link the two research areas by demonstrating that perisaccadic attentional dynamics improve predictions on scan path statistics. In a mathematical model, we integrated perisaccadic covert attention with dynamic scan path generation. Our model reproduces saccade amplitude distributions, angular statistics, intersaccadic turning angles, and their impact on fixation durations as well as inter-individual differences using Bayesian inference. Therefore, our result lend support to the relevance of perisaccadic attention to gaze statistics.

Published in **Nature Communications Biology, 2020**

2.1 Introduction

Visual perception in humans is the result of complex signal processing of visual input in the brain. Information enters the eyes at a rate of about 10^8 to 10^9 bit/s (Kelly, 1962). In order to handle this enormous amount of input, the visual system relies on foveation and selective attention (Yantis & Abrams, 2014). These two mechanisms reduce the information available at any given point in time to enable the brain to efficiently process the relevant aspects of visual information. *Foveation* refers to the decrease of visual acuity from the region extending about 2° around the point of fixation (the fovea) to the periphery of the visual field. During natural viewing, regions of interest are sequentially moved into the high resolution foveal area by saccadic eye movements (Findlay & Gilchrist, 2003; Henderson & Hollingworth, 2003). Natural vision is therefore an active process, determined by sequential choices of fixation locations. The resulting scan path (Noton & Stark, 1971a) is characterized by pronounced spatial correlations (Trukenbrod et al., 2019). *Selective attention* is the second key bottleneck of visual processing with a rate of about 100 bit/s (Zhaoping, 2014), prioritizing selected image regions at the cost of others. Under natural viewing conditions fixation position and visual attention are closely linked and coincide at the same location most of the time during viewing (Findlay & Gilchrist, 2003).

Experimentally, however, the locus of visual attention and fixation position can diverge, a condition referred to as covert attention (Posner, 1980; Posner & Cohen, 1984). Research on saccade dynamics in highly controlled experimental setups indicates that attention, as measured by processing benefits, precedes the fixation to the next saccade target (Deubel & Schneider, 1996; Hoffman & Subramaniam, 1995; Kowler & Blaser, 1995). Current models of eye movements and visual attention are typically based on the plausible simplification of directly equating location of attention and fixation position (Engbert, Trukenbrod, et al., 2015; Itti & Koch, 2001; Schütt et al., 2017; Tatler et al., 2017). Here we propose that perisaccadic covert attention shifts are an important factor in eye movement guidance.

The field of modeling eye movement behavior has primarily focused on predicting where fixations are placed in an image (Itti & Koch, 2000; Koch & Ullman, 1985; Kümmerer et al., 2017). The most advanced models are able to predict fixation density maps that closely resemble the empirical fixation densities they are based on (Bylinskii et al., 2015). The step from modeling static fixation densities to predicting scan paths reveals that bottom-up image information, while important, can not comprehensively explain the fixation selection process. This is illustrated by the fact that even a model that comprises no image information at all outperforms some static saliency models(Le Meur & Coutrot, 2016; Tatler & Vincent, 2009). Thus, scan path dynamics also play an important role. The ability of a model to predict human-like behavior can be much improved (Schütt et al., 2017) by adding basic dynamic mechanisms to the static image-based predictions (Le Meur & Liu, 2015; Rothkegel et al., 2016; Tatler et al., 2017; Tatler & Vincent, 2009).

Theoretical (Engbert, Trukenbrod, et al., 2015; Itti & Koch, 2001) and experimental work (Rothkegel et al., 2016) agree that two essential components in explaining dy-

namic scan paths are attentional selection and inhibitory tagging of previously fixated locations. The former refers to the combination of foveation and the attentional field, which defines a limited area from which information can be extracted. The attentional field is often represented as a Gaussian distribution, with its peak representing the fovea. Thus, as a first-order approximation, visual input is given by a Gaussian blob defined by the fixation position in a given scene. The second component keeps track of fixation history in order to drive exploration in scan paths and prevent continuous return to the same high-saliency regions (Klein, 2000). In behavioral experiments, *inhibition of return* has been widely found as a component of human visual behavior (Klein & MacInnes, 1999), electrophysiology (Hopfinger & Mangun, 1998), and, more recently, as a neural process in the frontal eye field (Mirpour et al., 2019).

Attentional selection and inhibitory tagging have been previously implemented in a dynamical model for scan path generation (Engbert, Trukenbrod, et al., 2015; Schütt et al., 2017). The SceneWalk model(Engbert, Trukenbrod, et al., 2015) serves as a platform for the current work on the analysis of the role of attention around the time of saccade. Conceptually the model comprises two independent streams, activation and inhibition, which are computed on discrete 128×128 grids mapped to the image dimensions. The activation stream is implemented as a Gaussian aperture around the current fixation location (see Eq. 2.1) convolved with a saliency map. This local saliency then evolves over time using a differential equation (see *Methods* for mathematical details), meaning that past fixations can influence the current activation stream. The inhibition stream implements fixation tagging by Gaussian maps centered around the fixation location and similarly evolving over time using a differential equation such that past fixations retain some influence over the current inhibition stream. The size of the Gaussian window $\sigma_{A/F}$, as well as the decay parameters $\omega_{A/F}$ and other free model parameters are jointly obtained from the parameter inference (see *Methods*). As illustrated in Fig. 2.1, activation and inhibition maps are subtractively combined to yield a priority map (Bisley & Mirpour, 2019), i.e., the 2D fixation probability map for the selection of the upcoming saccade target.

In the current context of perisaccadic processes, it is important to note that the strongest impact on mean fixation duration is generated by the variation in saccadic turning angles (Tatler et al., 2017). Continuing to move along the previous saccade's vector is associated with much shorter fixation durations than when the saccade direction changes by 90° or more (see the 80 ms effect in Fig. 2.6A). Therefore, we primarily seek to explain this coupling between fixation duration and saccade angle. Thus, we simplify our analysis by assuming random timing of fixation durations (assuming a gamma-distribution) and investigate the coupling with target selection under different turning angles. In future work the temporal control in the model could be extended to include other metrics (e.g., local saliency) for predicting fixation durations.

In this article we investigate a neurophysiologically plausible implementation of attentional dynamics and inhibitory principles. We extend the SceneWalk model (Engbert, Trukenbrod, et al., 2015) of eye-movement control by adding the concept of attentional shifts around the time of a saccade. Large-scale numerical simulations are carried out to estimate model parameters from experimental data using Bayesian data

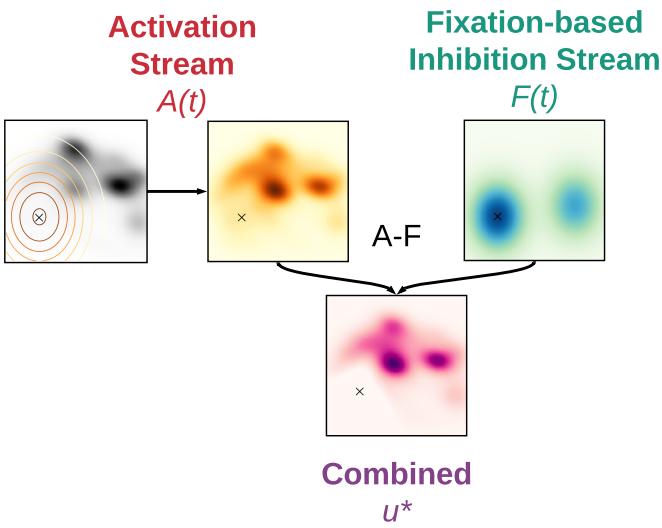

Figure 2.1 Attentional processing streams in a conceptual scan path model. Visual attention and inhibitory tagging are largely independent processing streams which evolve neural activations via time-dependent input and decay. Constraining a saliency map (black and white color map) by a Gaussian aperture can approximate the extent of visual attention (orange color maps), as shown on the left. Inhibitory tagging, shown in blue color maps, keeps track of previously visited locations, as shown on the right. The X marks the current fixation position. Combining the activation and inhibition streams yields a priority map from which fixation positions can be selected.

assimilation (Schütt et al., 2017). These covert perisaccadic attentional shifts turn out to improve model performance on a variety of eye movement statistics.

2.2 Results

The current work investigated the potential role of perisaccadic attention on human saccade statistics. In the next paragraph, we explain our theoretical model, before we describe experimental paradigm and experimental data.

2.2.1 Integrating perisaccadic attention with gaze control

Before the saccade is executed toward a target, performance benefits in accuracy and speed can be measured at the target location. This has frequently been interpreted as attention being allocated to the part of the image that is about to be fixated as part of saccadic planning. In Figure 2.2A (*leftmost*), we see that during a fixation, the fixation location and the center of attention are coaligned. Once the upcoming target location is selected from the priority map $u_{ij}(t)$ but before the saccade occurs (Fig. 2.2A, *second from left*), attention already moves to the upcoming saccade target, decoupling fixation (red 3-pointed star) and attention (green 5-pointed star). The concept that covert attention shifts precede saccadic eye movements is well-established in the literature (Deubel & Schneider, 1996; Irwin & Gordon, 1998), with clear evidence

for this *predictive attentional targeting* as early as 150 ms before saccade onset (Rolfs et al., 2011).

Furthermore, attention has been shown to move retinotopically with the saccade (Marino & Mazer, 2016). Thus, just after a saccade similar processing benefits can be found in a location along the saccade vector, which aligns with the retinotopic position of the target before the saccade (Golomb et al., 2008), a phenomenon called *retinotopic attentional trace* (RAT). The pre-allocated attention peak moves with the saccade such that it lands shifted along the saccade vector away from the saccade target. Figure 2.2A (*third from left*) shows that immediately after a saccade, attention is shifted to the same retinotopic position as the previous pre-saccadic shift and thus spatiotopically shifted in the same direction as the saccadic movement. Experimentally, the influence of the shift lasts about 100 to 200 ms (Golomb et al., 2008). After this interval the locus of activation moves to coincide with the fixation position again (Fig. 2.2A, *rightmost panel*). An alternative representation of the temporal progression of persaccadic processes in the model is available in the supplementary material (Fig. S2).

If we consider the added activation along the saccade vector as a component in saccade selection, this is in good agreement with the experimental finding of shorter fixation durations before forward saccades. The post-saccadic RAT is therefore the second part of the attentional decoupling that begins before saccade onset. Behavioral evidence for attentional shifts during a saccade (Deubel & Schneider, 1996) as well as neurophysiological correlates for post-saccadic retinotopic enhancements have been found (Golomb et al., 2010). Below we suggest that attentional shifts are a likely explanation for a systematic effect on saccade statistics observed during scan path formation. Figure 2.2B illustrates the influence of perisaccadic attentional shifts on the activation maps. The streams evolve over time (Eq. 2.4, 2.5). Each successive map consists of the previous map and the current new information in a ratio determined by the decay function. The model, thus, has infinite memory, although depending on the strength of the decay parameters, previous fixation's influence may decrease rapidly fixation targets are selected from the priority map (Eq. 2.6) at time $t_{fix} - \tau_{pre}$, where t_{fix} is the duration of the fixation and τ_{pre} is the duration of the pre-saccadic shift. Once the upcoming target is selected, attention moves to its location; after saccade execution, the post-saccadic attentional shift occurs; lastly, attention and fixation position are realigned when entering the main fixation phase (for details of the implementation see *Materials*).

In the experiments, 35 human observers viewed 30 natural color images (see *Materials*). We will compare simulations for the *baseline model* (Engbert, Trukenbrod, et al., 2015; Schütt et al., 2017) which includes only local saliency and inihibition evolving over time with the *extended model* that includes perisaccadic attention mechanisms. Model parameters for both models were estimated independently for each participant. For model fitting, fixation sequences of 2/3 of the images were used as training data, while all subsequent analyses were carried out on the remaining test images for each participant. The following section details some characteristic eye movement statistics found in experimental data.

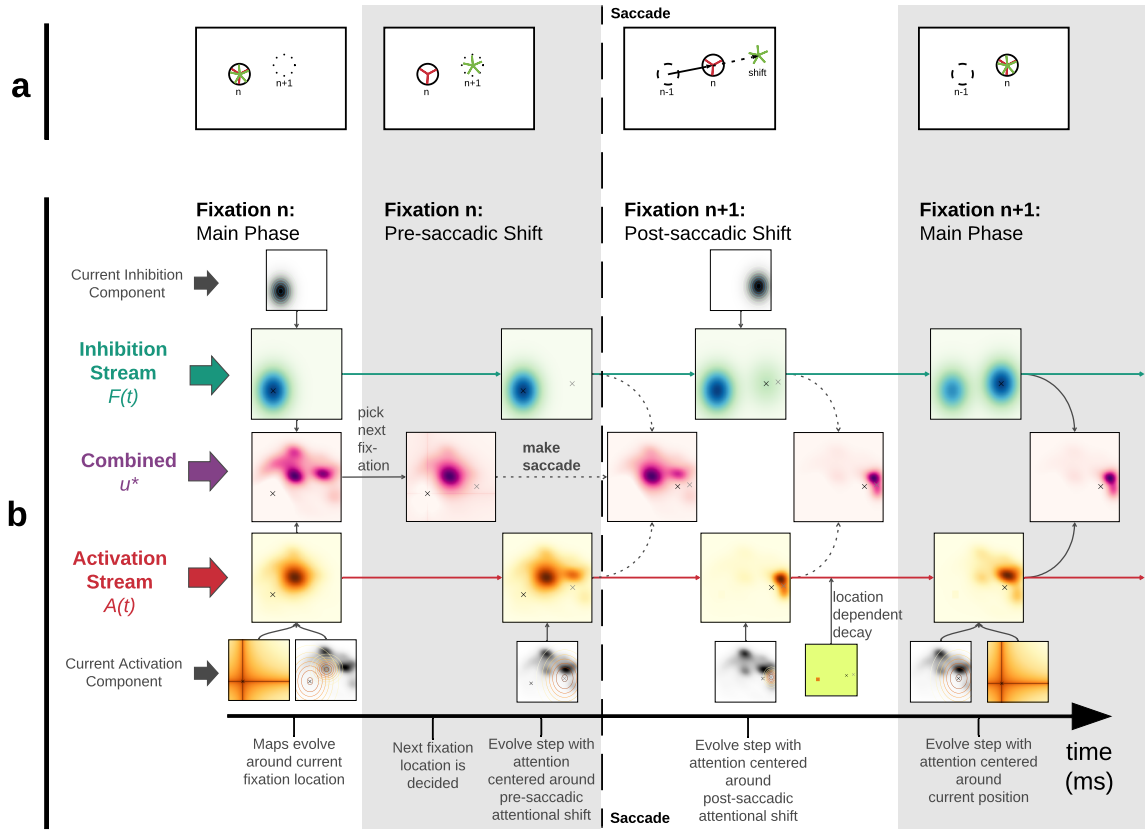

Figure 2.2 Timeline for processing around the time of saccade. (a) Attention and fixation position. *Leftmost panel:* During fixations, locus of attention (green 5-pointed star) and fixation position (red 3-pointed star) are aligned. *Second panel from left:* Immediately before a saccade, the upcoming fixation location has already been selected; attention moves to the target location (green), while fixation position remains at launch-site of the saccade. *Third from left:* After saccade execution, fixation position has been updated and, simultaneously, attention has shifted along the retinotopic activation trace (RAT) of the current fixation position before the saccade. *Rightmost:* During the fixation's main phase, the locus of attention and the fixation are realigned. **(b)** The activation (red-orange) and inhibition (blue-green) streams evolve over time during each of three model phases in each fixation. When a new fixation location is to be selected the streams are subtracted to yield a priority map (pink-purple). The activation map consists of a Gaussian aperture around a phase-dependent point in the image and image information as well as influences from past states of the model. The inhibition stream consists of a Gaussian aperture around the current fixation location and past states of the model.

2.2.2 Saccade amplitude distribution

The distribution of saccade amplitudes generated during a scene viewing experiment varies across participants and images. Overall, both the baseline and the extended models reproduce the qualitative shape of the saccade amplitude distribution (Fig. 2.3A) (Bahill et al., 1975; Bruce & Tsotsos, 2009; Tatler et al., 2011; Tatler, Vincent, et al., 2008). The experimentally observed saccade amplitude distribution is right-skewed, reflecting that amplitudes tend to be smaller than computer-generated saccades obtained by random sampling from the static 2D fixation density (Engbert, Trukenbrod, et al., 2015; Trukenbrod et al., 2019). Previously, we suggested this drop in saccade amplitudes is caused by the foveated visual system, which preferentially selects saccade targets from within attentional span. Therefore, inter-individual differences in mean saccade amplitudes should correlate with the size of the attentional span σ_A, which is defined as the standard deviation parameter of the Gaussian-shaped attentional blob (see Eq. 2.1). In Figure 2.3B, we show the expected correlation between σ_A and mean of saccade amplitude across participants, indicating that a larger area does indeed lead to longer saccades.

This statistic is perhaps the most prominent and intuitive. Previous modeling studies, like our baseline model, have been able to capture it as well as the extended model. The result we show here confirms that our addition of more complex mechanisms has not come at the cost of the more basic effects.

Additionally the improved fitting procedure allows both models to be fit separately for each subject. With model parameters estimated for each participant using the training images, the predicted mean saccade amplitudes for test images were compared to experimentally observed mean saccade amplitudes. We found good agreement between predicted and experimentally observed mean saccade amplitudes (Fig. 2.3C) indicated by a high correlation ($r = 0.91$). Our model is able to explain the inter-individual differences in the data via parameter variation.

2.2.3 Absolute and relative saccade angle distributions

Saccade angles are another important characteristic of human eye movement behavior. The absolute angle distribution reports the directions of saccades relative to the image frame. Interestingly, there is a strongly image-dependent tendency, which varies mostly with the distribution of image features. On average the distribution shows characteristic peaks in the four cardinal directions (Foulsham et al., 2008; Gilchrist & Harvey, 2006). Figure 2.4A shows that, the baseline model does not show the pronounced pattern found in experimental data. Comparatively the extended model shows a clear improvement with distinct peaks at 0°, 90°, 180°, 270°, and 360°. The extended model implements a mechanism for an oculomotor potential (see Equations 2.14 and 2.15), which preferentially weights the activation in the cardinal directions (Engbert et al., 2011) before it is combined with the inhibition stream.

The saccade turning angle distribution characterizes the relationship of consecutive saccades. In the experimental data there is a clear bias towards forward saccades,

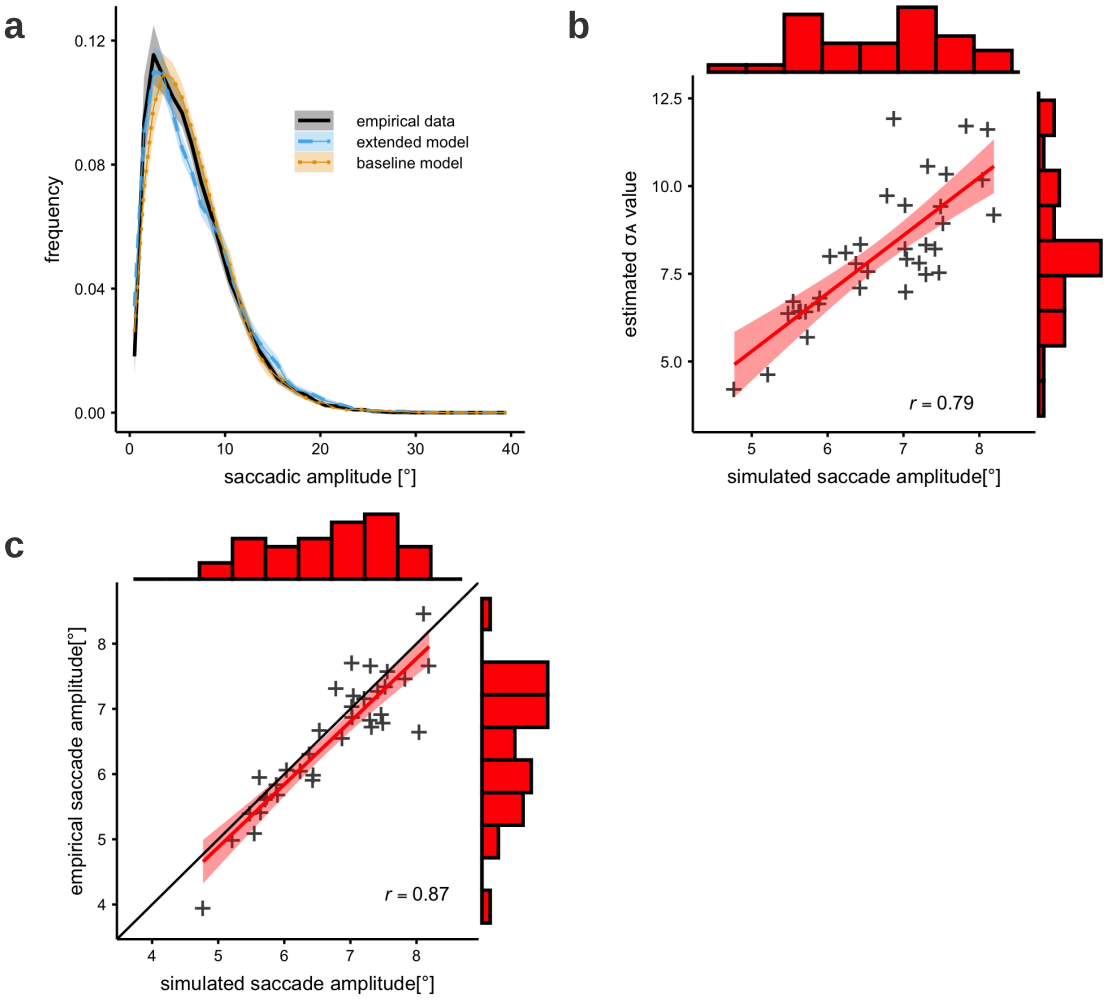

Figure 2.3 Saccade amplitude distribution and inter-individual differences. (a) The saccade amplitude distribution for experimental data (black line), the baseline SceneWalk model (blue, dashed line) and the extended model (yellow, dotted line). Shading represents the 95% confidence interval between subjects. **(b)** Size parameter σ_A of the attentional span and is positively correlated simulated mean saccade amplitude. **(c)** A high correlation is observed between experimental and simulated data on test images, using parameter estimates for each participant obtained for training data.

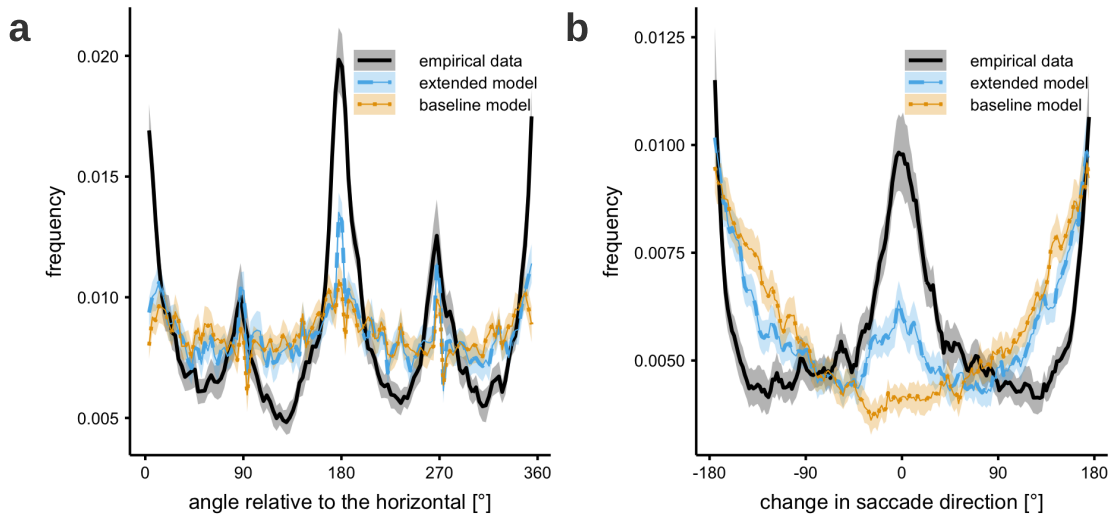

Figure 2.4 Saccade angle distributions. (a) Absolute angle distribution in the empirical and simulated data. Empirical data show a strong tendency for saccading in the cardinal directions. This is strongly image dependent and not specifically considered in the models. The shading shows the 95% confidence interval between subjects. **(b)** Saccadic turning angle distribution in the empirical and simulated data. The angle shown is the divergence from the previous saccade direction. Empirically we find a tendency to continue along the previous saccade vector or completely reverse it. The extended model partly shows this behavior. The shading shows the 95% confidence interval between subjects.

which follow the same vector of motion, and a secondary preference for return saccades, which reverse the saccade vector. Therefore, we should expect clear peaks at 0° and 180° in the corresponding turning angle distribution (Rothkegel et al., 2018; Rothkegel et al., 2016; T. J. Smith & Henderson, 2009; Tatler & Vincent, 2009). Figure 2.4B shows the results of the baseline and extended models in comparison to experimental data. The baseline model produces a u-shaped distribution without any indication of a forward bias. There is an increased probability of turning by about 180°, since the edges of the image represent hard constraints. This effect is large enough to overshadow the effects of return saccades that directly return to the previous fixation location (of which there are comparatively few). The extended model does develop a peak for forward saccades, showing better qualitative agreement with the experimental data, although the bias towards forward saccades is clearly weaker than in the experiment. The model's slightly muted responses could be caused by a number of factors, not least of which is the fact that the chosen general purpose likelihood procedure does not specifically target this metric. The indirect fitting of parameters supports the existence of the directional biases but may capture them only partially in the presence of other variance in the data.

The statistical preference of observers to maintain current saccade direction has been referred to as saccadic momentum (Luke et al., 2014; Rothkegel et al., 2018; T. J. Smith & Henderson, 2009; Wilming et al., 2013). Here we propose that the

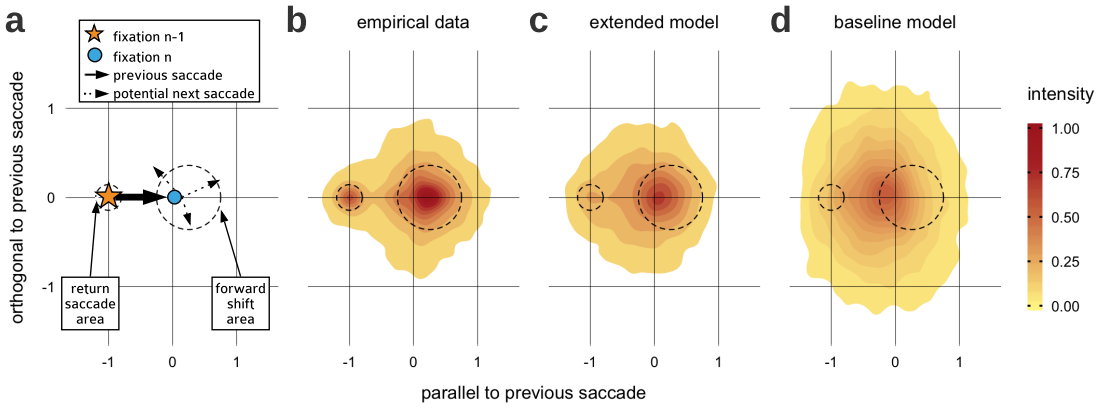

Figure 2.5 Joint probability of saccade turning angle and saccade amplitude normalized to the previous saccade. (a) Legend for the joint probability plot. The coordinate system is normalized relative to the previous saccade. **(b)** The experimental probability shows the return and forward peaks. **(c)** The extended model captures these characteristic properties qualitatively. **(D)** In the baseline model, neither the return peak nor the forward peak can be found.

experimental effect is at least partially due to attentional enhancement in the current saccade direction, which generates a peak in the attention map that produced the forward bias.

2.2.4 Joint probability of intersaccadic angle and amplitude

More generally, we can identify potential dependencies of saccade turning angle and saccade amplitude by visualizing the corresponding joint probability (Fig. 2.5). As discussed above, compared to all other directions, there is a pronounced tendency for saccades to either maintain or completely reverse the direction of the previous saccade. This effect is well documented in the literature (Rothkegel et al., 2018; Rothkegel et al., 2016; T. J. Smith & Henderson, 2009; Tatler, Vincent, et al., 2008; Tatler & Vincent, 2009) and is independent of a variety of other factors such as image content. The values on the axes in Figure 2.5 are relative to direction and amplitude of the previous saccade. In this normalized coordinate space, the previous saccade moved from position $(-1, 0)$ to position $(0, 0)$. The plotted density indicates the probability of the following saccade to be executed in a direction and with an amplitude relative to the previous saccade. Figure 2.5B reveals that there are two clear peaks in the experimental data, i.e., the return peak to the normalized launch site $(-1, 0)$ of the previous saccade and the forward peak that is related to the saccadic momentum effect discussed above. It is important to note that the experimental return peak is not particularly high, but it is distinct since surrounding 2D regions do not exhibit a high fixation density.

In our extended model (Fig. 2.5C), the mechanism responsible for the forward saccades is the attentional shift before and after a saccade (Eq. 2.9 - 2.11). The distinctive

shape of the return saccade peak, we suggest, is the result of the combination of a slow, global inhibition of return and a directed smaller facilitation of return (Eq. 2.12) (see *Materials*). The former is implemented as the model's inhibition stream, while the latter is implemented as reduction in decay speed in the attention map, localized at the previous fixation location. The baseline model cannot produce the return and forward peak, since it lacks the mechanistic principles for coupling subsequent saccades.

2.2.5 Intersaccadic angle and fixation duration and saccadic amplitude

The next two analyses correspond to the interdependence of fixation duration and saccade amplitude, and saccadic turning angles. Both have a distinctive shape in the data, showing that forward saccades tend to be shorter and preceded by shorter fixations, while changing direction takes longer and evokes longer saccades. Pilot simulations indicated that the effect reported in this section are not due to the addition of the oculomotor potential.

The new model notably improves the fit of the dependence of fixation duration on the turning angles (see Fig. 2.6). While previously there was no temporal component in the model the added phases of shifted activation enable the model to dynamically respond to the duration of a fixation. In the model, each fixation begins with the post-saccadic shift phase. In terms of the attention activation map, this means that there is more activation along the previous saccade vector. After this phase the influence of the shift diminishes. Thus, when the fixation is short, there is still a lot of influence from the shift, increasing the chance of producing a forward saccade. When the fixation is long, the influence of the post saccadic shift has subsided, allowing for activation from other salient locations to guide the saccade.

2.2.6 Likelihood-based comparison

Since our approach includes the likelihood computation of the baseline and extended models, we can make use of the models' likelihood functions for model comparison (Schütt et al., 2017). This approach entails evaluating the model likelihood given the empirical test data and computing the average log-likelihood per fixation of all scan paths. We then compare this metric to previous models (Kümmerer et al., 2015).

The overall likelihood of the model given the data is larger for the extended model than for the original model (Fig. 2.7). In general, improved likelihood indicates improved predictive power of a model. The additions to the baseline model discussed in the current study, though theoretically well-founded, were extensive and considerably increased the model complexity. Conceivably adding these mechanisms could have led to improved scan path dynamics but worsened overall likelihood predictions, or else made the model volatile or unstable. In general, the likelihood is an objective measure of overall model performance(Schütt et al., 2017). As we have seen, the extended model performs much better than the baseline model at a number of qualitative eye-movement effects, while the improvement in general model likelihood is relatively small. Effects such as the impact of saccade turning angles on saccade amplitude

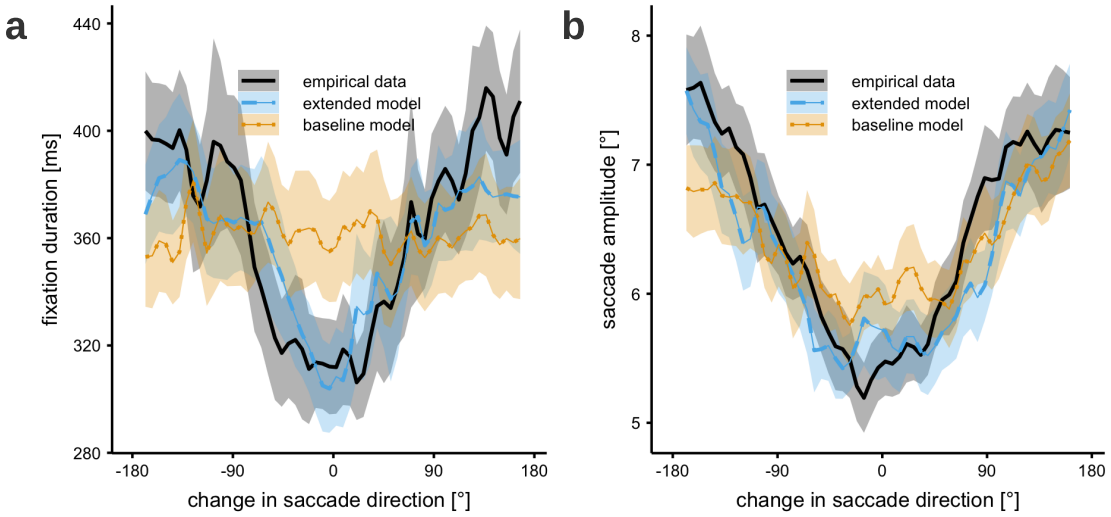

Figure 2.6 Average saccade amplitude and fixation duration are related to the change in saccade direction (saccadic turning angle). **(a)** Fixation duration is shortest for saccade moving forward. Results from the extended model are in good agreement with experimental data. **(b)** Saccade amplitude is smallest for forward saccades and largest for return saccades. While the baseline model reproduces this effect qualitatively, the extended model produces a better fit to the experimental data.

are strong and important for biological plausibility of the model. At the same time, however, the impact on the overall likelihood is limited, since their contribution to 2D fixation density is small. In combination, the large improvements in eye-movement statistics and relative improvements in likelihood across model variants allow a strong conclusion in favor of the proposed model extension.

2.3 Discussion

Moving from models of static fixation probabilities to the generation of scan paths has recently begun to attract interest in the field of attention modeling (Engbert, Trukenbrod, et al., 2015; Le Meur & Liu, 2015; Schütt et al., 2017; Tatler et al., 2017; Zelinsky, 2008). The success of saliency-based visual attention modeling (Itti & Koch, 2001; Kümmerer et al., 2017; Kümmerer, Wallis, et al., 2014; Schütt et al., 2019) over the last 30 years makes a strong case for the use of priority maps (Bisley & Mirpour, 2019) as a core component in scan path generation. In addition to image and task influences biologically represented in priority maps, scan paths on scenes are also characterized by a number of statistical characteristics, e.g., saccade angles and modulations of fixation duration or saccade amplitude by saccadic turning angles. Our modeling study lends support to the fact that attentional dynamics around the time of saccade exert a fundamental influence on the behavioral statistics of scan paths.

Previous research on visual attention shows that processing resources are covertly al-

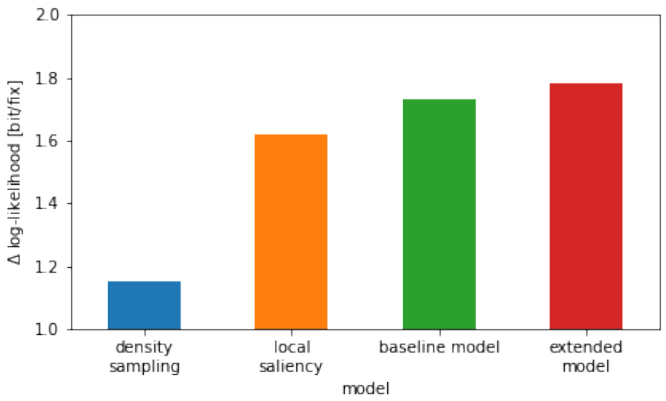

Figure 2.7 General model likelihood of models fitted on the training data given the test data. Density sampling draws fixations directly from the empirical fixation distribution without dynamics. Local saliency produces scan paths by picking from the fixation density filtered through a Gaussian window, with no dynamics. The baseline model and the extended model are the dynamic models described in this article.

located away from the current fixation location just before (Deubel & Schneider, 1996; Irwin & Gordon, 1998; Rolfs et al., 2011) and just after (Golomb et al., 2008; Golomb et al., 2010; Rolfs et al., 2011) a saccade is produced. In this study, we added shifts of covert attention to a dynamical model of scan path generation (Engbert, Trukenbrod, et al., 2015; Schütt et al., 2017) and find improved agreement with gaze statistics observed in experimental data. Most importantly, the characteristic distribution of saccadic turning angles with a clear bias towards forward and return saccades and the influences of saccadic turning angle on fixation durations and saccade amplitudes can be explained partly by covert attention shifts around the time of a saccade. The importance of covert attention and perisaccadic mechanisms is apparent throughout the visual system, both at the macroscopic as well as at the microsaccade levels (Engbert, 2012; Tian et al., 2016, 2018).

The first generation of computational models in scene viewing were static models which predicted fixation locations on any given image based on statistical image features. The strength of these static models lies in producing densities that resemble empirical fixation density maps. Recently, the predictive power of some models has become close to perfect and approached the gold standard (Kümmerer et al., 2017; Kümmerer et al., 2015). However, by design these models do not take temporal dynamics within a scan path and the inhomogeneity of the retinal acuity into account. From this perspective, it is not surprising that static models predict fixation density, but not sequences of fixations (Foulsham & Underwood, 2008; Rothkegel et al., 2016; Schütt et al., 2017). This simple fact points to the interesting observation that eye movements in scene viewing are guided in large part, but not exclusively by observer-

and image-specific factors. Human eye movements are influenced by oculomotor and attention systems, producing pervasive systematic statistical tendencies in experimental data.

Previously published dynamic models outperform static models substantially (Le Meur & Liu, 2015; Schütt et al., 2017). The most evident feature of the human visual system which indisputably influences scan path dynamics is foveation. Accordingly, even a minimal model like weighting a saliency map by the distance to a current fixation location significantly improves model performance (Parkhurst et al., 2002). The SceneWalk model (Engbert, Trukenbrod, et al., 2015), which served as a baseline for our study, incorporates foveated saliency in its activation stream. A further advance in the modeling of scan paths has been the addition of inhibitory fixation tagging (Itti et al., 1998; Klein, 2000; Posner et al., 1985). The baseline model implements such an inhibition stream as a second component shaping the priority map (Bisley & Mirpour, 2019) by difference of activation.

The fact that long fixations often occur in frequently fixated areas (Einhäuser & Nuthmann, 2016) implies that fixation duration and target selection are related. The LATEST model (Tatler et al., 2017) combines the prediction of scan paths and fixation durations by interpreting scan paths as a continuous series of stay (maintain fixation) or go (saccade) decision (Carpenter & Reddi, 2001; Ratcliff & McKoon, 2008; Reddi & Carpenter, 2000). Each individual location on a weighted saliency map influences two LATER units (Noorani & Carpenter, 2016), i.e., one for normal and long latencies and one for short latencies. These units accumulate evidence from each location in the image until one reaches a threshold depending on the current location, triggering a saccade. The accumulation rate of each location in the image is controlled by image-content factors like image features and semantic interest, as well as by oculomotor factors like the change in saccade direction and target eccentricity. Coupling of experimental data and model is achieved by statistical linear mixed-effects modeling. Thus, the LATEST model makes little attempt at explaining the origin of the factors that influence the rate of evidence accumulation, instead focusing on the specific selection mechanism and its relationship with fixation duration. By contrast, the extended SceneWalk model is based on mechanistic assumptions derived from neural and cognitive knowledge about the contributing factors to fixation selection. Parameters are based on statistically rigorous likelihood approach that evaluates the model assumptions given the data.

Generally, the value of a model must be quantified in terms of predictive power and explanatory value. For the models discussed here, we carried out comparisons of simulated scan paths and human eye movement data. A number of metrics have been proposed for such a comparison (Cerf et al., 2008; Jarodzka et al., 2010; Mannan et al., 1996). Critically, however, the choice of individual statistics has a crucial influence on the outcome and there is, in most cases, no rigorous justification for the used metric. A solution to this is to evaluate dynamical scan path models using a likelihood approach (Schütt et al., 2017), which provides a statistically well-founded and reliable measure for the predictive quality of a dynamical model. In this article we relied on Bayesian data assimilation (Reich & Cotter, 2015) as a statistically rigorous framework

for testing whether the model architecture accurately represents the data generation process. This approach turned out to be particularly fruitful for strongly theory-guided models. Using general likelihood to estimate parameters of the model lends credibility to the theoretical foundations from eye movement literature implemented by the model.

In addition to better predicting human scan paths during scenes viewing, the integration of biologically-inspired attentional dynamics into models of eye guidance unifies two very disparate fields of eye movement research. The research into covert attention shifts and perisaccadic effects is typically concerned with processes that occur on a highly detailed level in very controlled experimental setups. By contrast scene viewing literature usually operates at a higher level, on which the minutia of saccade programming or covert attention are typically passed over. Thus, influences arising from the microscopic level of eye movement control can explain effects we observe at the macroscopic level.

2.4 Methods

Experiment Experimental data for this study were collected in a larger corpus study on scene viewing which is described in detail elsewhere (Rothkegel, Schütt, et al., 2019; Schütt et al., 2019). Images and fixation data from this corpus experiment can be downloaded from an Open Science Foundation repository (see below(Rothkegel, Schütt, et al., 2019)). The corpus consists of eye movement data from 105 participants viewing 90 images of natural or urban landscapes from 6 different categories for a fixed duration (10 s). Each category contained 15 images. Images were chosen such that the most interesting image parts either fell on the left, right, upper, lower or, central image side (Fig. S1 provides some examples). The last category were images with natural patterns, minimizing the influence of particularly salient objects. During the viewing subjects were given no task except to freely view the images.

In this study we used Experiment 3 from the corpus study (Rothkegel, Schütt, et al., 2019; Schütt et al., 2019), in which participants viewed color images. This subset of data contains the eye movements of 35 participants, who viewed 30 images from each category without a task. We further split the data set into test and training data by randomly choosing 1/3 of the images (10 from each category) for each participant.

For saccade detection we applied a velocity-based algorithm (Engbert & Mergenthaler, 2006; Engbert & Kliegl, 2003). Saccades had a minimum amplitude of 0.5° and exceeded an average velocity during a trial by six (median-based) standard deviations for at least six data samples (12 ms). The epoch between two subsequent saccades was defined as a fixation. After preparation, 312, 267 fixations and saccades were detected for further analysis.

Baseline Model The original SceneWalk model (Engbert, Trukenbrod, et al., 2015) was implemented on a 128×128 grid, where (x, y) give the physical coordinates in degrees. For each fixation in the scan path we start by computing simple 2-D Gaussians

centered at current fixation position (x_f, y_f) for both the inhibition and the attention pathway, each with an appropriate standard deviation $\sigma_{A/F}$ (A denotes the attention stream, F denotes the fixation stream to generate inhibitory tagging).

$$G_{A/F}(x,y) = \frac{1}{2\pi\sigma_{A/F}^2} \exp\left(-\frac{(x-x_f)^2+(y-y_f)^2}{2\sigma_{A/F}^2}\right) \tag{2.1}$$

Both the inhibition $F_{ij}(t)$ and the activation $A_{ij}(t)$ streams evolve over time under current visual input and decay (due to limited of visual memory), i.e.,

$$\frac{\mathrm{d}A_{ij}(t)}{\mathrm{d}t} = \omega_A\left(\frac{S_{ij}\,G_A(x_i,y_j;x_f,y_f)}{\sum_{kl} S_{kl}\,G_A(x_k,y_l;x_f,y_f)} - A_{ij}(t)\right) \tag{2.2}$$

$$\frac{\mathrm{d}F_{ij}(t)}{\mathrm{d}t} = \omega_F\left(\frac{G_F(x_i,y_j;x_f,y_f)}{\sum_{kl} G_F(x_k,y_l;x_f,y_f)} - F_{ij}(t)\right), \tag{2.3}$$

where the input to the activation maps is the Gaussian-weighted local saliency $S_{kl}G_A(x_k,y_l;x_f,y_f)$ and the input to the inhibition map is a Gaussian blob at current fixation position.

The differential equations that determine the temporal evolution of the activation maps, Eq. 2.2 for the attention map and Eq. 2.3 for the fixation/inhibition map, can be integrated analytically to provide a closed solution for the activation changes during fixation, i.e.,

$$A(t) = \frac{G_A S}{\sum G_A S} + e^{-\omega_A(t-t_0)}\left(A_0 - \frac{G_A S}{\sum G_A S}\right), \tag{2.4}$$

and

$$F(t) = \frac{G_F}{\sum G_F} + e^{-\omega_F(t-t_0)}\left(F_0 - \frac{G_F}{\sum G_F}\right), \tag{2.5}$$

where we dropped the indices i, j for simplicity. In the equations, the term $e^{-\omega_{A/F}(t-t_0)}$ determines the speed of decay of the past states of the map.

Next, both activation maps were combined to compute the priority map $u_{ij}(t)$,

$$u_{ij}(t) = \frac{(A_{ij}(t))^\gamma}{\sum_{kl}(A_{kl}(t))^\gamma} - C_F\frac{(F_{ij}(t))^\gamma}{\sum_{kl}(F_{kl}(t))^\gamma}\,. \tag{2.6}$$

Mathematically, the two maps are shaped by exponent γ before subtraction, and a weight parameter C_F for inhibition is introduced. We expect $\gamma \approx 1$, equivalent to Luce's choice rule (Luce & Raiffa, 1989).

As subtraction can cause negative activation, in the next step we take only the positive component of the map,

$$u^*_{ij}(u_{ij}) = \begin{array}{ll} u_{ij}, & \text{if } u_{ij} > 0 \\ 0, & \text{otherwise} \end{array} \tag{2.7}$$

Phase	Start	End	F center	A center
Post-saccadic shift	0	τ_{post}	$\mathrm{f}ix_n$	$\mathrm{r}emap$
Main (no shift)	τ_{post}	$t_{fix} - \tau_{pre}$	$\mathrm{f}ix_n$	$\mathrm{f}ix_n$
Pre-saccadic shift	$t_{fix} - \tau_{pre}$	t_{fix}	$\mathrm{f}ix_n$	$\mathrm{f}ix_{n+1}$

Table 2.1 Model Phases: onset times and locations around which the Gaussians in both streams are centered. Parameter t_{fix} indicates the fixation's duration, parameters τ_{pre}, τ_{post} are the phase durations, and parameters fix_{n+1}, fix_n, and $\mathrm{r}emap$ are the locations.

and, finally, add noise ζ

$$\pi(i,j) = (1-\zeta)\frac{u^*_{ij}}{\sum_{kl} u^*_{kl}} + \zeta\frac{1}{\sum_{kl} 1} \tag{2.8}$$

to obtain the probability map $\pi(i,j)$ for the selection of saccade targets. This process is repeated for each fixation in a sequence, where the current state information is combined with the past activation maps to produce a continuously evolving prediction of the next fixation.

The model structure reveals the following parameters: (1, 2) σ_A and σ_F, which are the standard deviations of the current fixation's attention and inhibition Gaussians respectively, (3, 4) ω_A and ω_F, which are the speed at which past states of the model lose influence over the current, (5) γ, the shaping parameter for the Gaussians, (6) the coupling factor C_F, which is the weight of the inhibition pathway, and (7) the noise parameter ζ determining the background noise for the probability map $\pi(i,j)$.

Pre-saccadic attentional shifts Once a new fixation location is chosen the center of attention moves to the upcoming fixation location, while the center of the inhibition map remains at the current fixation location (see Table 2.1). In the model, the pre-saccadic shift is implemented by moving the attentional Gaussian to center around the next fixation location, while the inhibition remains in the same position for a time τ_{pre}. The inhibition stream is calculated for the entire fixation duration using Eq. 2.5, therefore, we have

$$G_A^{\mathrm{pre}}(x,y) = \frac{1}{2\pi\sigma_A^2}\exp\left(-\frac{(x-x_{f+1})^2+(y-y_{f+1})^2}{2\sigma_A^2}\right), \tag{2.9}$$

and then continue computations using Eqs. 2.4-2.5 with G_A^{pre} instead of G_A for the duration of τ_{pre}. When the pre-saccadic phase terminates, the saccade is executed.

Post-saccadic attentional shifts The center of the post-saccade attention peak is determined by extending the vector of the preceding saccade by a shift amplitude η, i.e.,

$$(x_{\mathrm{s}}, y_{\mathrm{s}}) = (x_n, y_n) + \frac{(x_\delta, y_\delta)}{\sqrt{x_\delta^2 + y_\delta^2}} \cdot \eta\,, \tag{2.10}$$

where the saccade direction is given by the vector (x_δ, y_δ) with $x_\delta = x_n - x_{n-1}$ and $y_\delta = y_n - y_{n-1}$. Thus, the attentional Gaussian is centered at the shifted location

$$G_A^{post}(x,y) = \frac{1}{2\pi\sigma_{post}^2} \exp\left(-\frac{(x-x_\mathrm{s})^2 + (y-y_\mathrm{s})^2}{2\sigma_{post}^2}\right) \tag{2.11}$$

during the post-saccadic phase. Temporal evolution of activation maps continues based on Eqs. 2.4-2.5 with G_A^{post} instead of G_A for a duration of τ_{post}. Meanwhile the inhibition stream evolves with the center of inhibition in the same location as the fixation position.

After the post-saccadic shift phase, the cycle is completed and another main phase follows. The attention center moves to each of the three locations in turn via discrete steps as shown in Table 2.1. We have chosen this discrete approximation with constant durations of pre- and post-saccadic shifts to compute activation changes in all fixation phase efficiently. Neurophysiological support for our discrete approximation has been found (Golomb et al., 2010), indicating that attention does not move smoothly over space from location n to location $n+1$ but instead selectively starts building up at the target location $n+1$.

Facilitation of Return To account for Facilitation of Return (FoR) we implement a selectively slower decay of the attention map in a spatial window centered at the previous fixation location. Different from the overall decay rate ω_A, we define a reduced decay rate $\omega_{\mathrm{F}oR}$ for a window $x-\nu < x_{f-1} < x+\nu$ and $y-\nu < y_{f-1} < y+\nu$ around the previous fixation location (x_{f-1}, y_{f-1}), where ν is the size of the window. Therefore, reduced decay of activation in the attention map, Eq. 2.4, is given by

$$A(t) = \frac{G_A S}{\sum G_A S} + e^{-\omega_{FoR}(t-t_0)}\left(A_0 - \frac{G_A S}{\sum G_A S}\right) \tag{2.12}$$

for the spatial window defined above.

In addition to the strongly attention-related mechanisms above, we added the following two less dynamic and more general biases.

Center Bias The original SceneWalk model initiates its activation maps with uniform distributions. While it is difficult accurately know the initial state of the visual system when viewing images, previous work has shown that the central fixation bias has a strong influence on the first fixation. Starting the model with a central activation improves the predictions of the model (Rothkegel et al., 2017). In line with this finding, we also initiated the model with central activation. The evolution equation for the first fixation is

$$A(t) = \frac{G_{fix} S}{\sum G_{fix} S} + e^{-\omega_{cb}(t-t_0)}\left(A_{0_{CB}} - \frac{G_{fix} S}{\sum G_{fix} S}\right) \tag{2.13}$$

Oculomotor Potential Research into the oculomotor system has revealed a marked preference for saccades in the cardinal directions. In order to implement this tendency in the model we introduced an additive occulomotor component. A plus-shaped oculomotor map centered on the current fixation position is generated

$$OMP = \left((x - x_f)^2 \cdot (y - y_f)^2\right)^{\chi}, \tag{2.14}$$

where the factor χ determines the steepness of the slopes. The oculomotor map is added to the combined map u_{ij}, before the normalization and the addition of noise (Eq. 2.7, 2.8)

$$u_{OMP} = u + \left(\psi \cdot \left| -\frac{OMP}{\sum(OMP)} \right|\right). \tag{2.15}$$

Additional model parameters The implementation of the extended SceneWalk model gives rise to several new parameters. To the 7 parameters of the original model we add (a) ω_{cb}, the decay speed of the center bias, (b, c) σ_{cb_x} and σ_{cb_y}, the size of the center bias, (d, e) τ_{pre} and τ_{post}, the durations of the attention shift phases, (f) η, the distance of the post-saccadic shift, (g) σ_{post} the size of the shifted Gaussian, (h, i) ω_{FoR}, the attention decay at the previous fixation position and ν, the size of the facilitation window, and (j, k) the steepness χ and factor ψ of the oculomotor potential.

Estimated and fixed model parameters We implemented a fully Bayesian approach to parameter inference (Schütt et al., 2017) using numerical computation of the models' likelihood functions and advanced Monte Chain Monte Carlo (MCMC) techniques. Details are given in the next section. For a discussion of the full results including marginal posterior densities see below (Section entitled *Detailed results on parameter estimation*).

In Table 2.2 we report point estimates for all parameters as averages over participants. The full estimates for each participant can be found in the *Supplementary Materials (Table S2 and Table S3*. These point estimates were computed from the posterior densities by determining the highest posterior density region for an alpha of 0.5 (i.e., the highest 50% of the density are in this region), assuming a unimodal distribution. The reported credibility intervals the lower and upper bounds of the highest density interval. The point estimate for the parameter represents the center of the highest posterior density interval.

Some of the model parameters could be constrained by the physiological literature and some of the parameters had to fixed in order to improve convergence of the parameter estimation. The latter case was checked by large-scale pilot simulations with different model versions using a separate data set. In Table S1 we list all fixed model parameters.

First, we separated the time scales of attention and inhibition stream by one order of magnitude, i.e., $\omega_F = \omega_A/10$. We assume ω_F is slower to decay by a magnitude than ω_A, to enable long term inhibition of return and fast build-up of activation for

Parameter	Baseline SW Mean	Baseline SW +/-	Extended SW Mean	Extended SW +/-
ω_A	14.802	2.555	9.996	2.391
σ_A	7.482	1.165	7.320	1.004
σ_F	4.629	1.041	6.834	2.626
γ	0.935	0.095	0.956	0.102
$log(\zeta)$	-1.132	0.131	-1.727	0.260
χ	-	-	0.059	0.028
η	-	-	0.415	0.105
$log(\psi)$	-	-	-0.613	0.192

Table 2.2 Maximum posterior density estimates of the model parameter estimations of all subjects and credibility intervals (see text).

attentional capture. Second, we set $C_F = 0.3$, where the numerical value was obtained from pilot simulations indicating that the relative influence of the inhibition stream must be smaller (but not negligible) compared to the corresponding influence of the attention stream.

In the extended model, some of the additional parameters need further discussion. First, we set $\sigma_{CB} = 4.3$ and $\omega_{CB} = 1.5$ as described in (Rothkegel et al., 2017), for a typically sized center bias and an attention decay that is slower than the regular ω. The center bias parameters are difficult to estimate, since their influence is mainly limited to the first fixation. Second, we fixed $\omega_{FoR} = \omega_A/10$, representing an approximate value for decay slower by a magnitude and the size of the facilitation of return window to be approximately the size of the fovea, i.e., 2° of visual angle. As before, only a relatively small amount of fixations are influenced by this mechanism, making it difficult to identify the numerical value reliably. Third, we set the times for post- and pre-saccadic attentional shifts to $\tau_{pre} = 0.1$ s and $\tau_{post} = 0.05$ s, where the numerical values are determined by pilot simulations. Due to their small magnitude, values for ζ and ψ were estimated in the log scale.

Bayesian parameter inference Parameter inference of the dynamical models discussed here was implemented in the general framework of data assimilation (Reich & Cotter, 2015) using a fully Bayesian estimation procedure (Rabe et al., 2021; Schütt et al., 2017; Seelig et al., 2020). In this statistical inference we used the computation of the models' likelihood functions. Given a fixation sequence $f_1 \dots f_{i-1}$, where each fixation f_i is determined by its coordinates $f_i = (x_i, y_i)$, the likelihood of the model specified by a set of parameters θ can be computed as a product of probabilities, i.e.,

$$L_M(\theta|\mathrm{d}ata) = P_M(f_1) \cdot \prod_{i=2}^{n} P_M(f_i|f_1, \dots, f_{i-1}, \theta) \,, \tag{2.16}$$

where $P_M(f_1)$ is the probability of the initial fixation starting at time $t = 0$ and the conditional probabilities $P_M(f_i|f_1 \dots f_{i-1}, \theta)$ can be read off from the models priority map $\pi(i, j)$.

For scaling and numerical reasons the log-likelihood is usually used. Thus, the sum of the scan path's log-likelihood per fixation for the entire data set gives one value that characterizes model performance. As suggested by (Schütt et al., 2017), taking the log_2 of the likelihood enables the use of the unit bit. A null model, in which the probability of choosing each point a 128×128 pixel image is the constant, would be $\log_2(1/128^2) = -14$. A hypothetical model which, unrealistically, perfectly predicts the data would have a log-likelihood of 0. It is important to note that for model comparison we can take the mean log-likelihood per fixation while for the parameter estimation the non-normalized sum log-likelihood of a scan path is the appropriate measure.

Based on the likelihood $L_M(\theta|\mathrm{d}ata)$ and a prior distribution $P(\theta)$, the posterior distribution is computed via Bayes' rule as

$$P(\theta|\mathrm{d}ata) = \frac{L(\theta|\mathrm{d}ata)P(\theta)}{\int_\Omega P(\theta)L(\theta|\mathrm{d}ata)d\theta}\,, \tag{2.17}$$

where typically a Markov Chain Monte Carlo (MCMC) approach is needed to compute the posterior density numerically. For our parameter estimations we used the implementation of the DREAM Algorithm that is published as PyDream (Laloy & Vrugt, 2012). Each estimation ran three chains of 20,000 iterations. Since the DREAM estimation procedure requires a large number of model evaluations, the computing time of the likelihood function in critical for the baseline SceneWalk model and, in particular, for the extended SceneWalk model. We therefore implemented parallel computations of the likelihood for fixation sequences. The priors, loosely based on pilot estimations on a separate dataset, were chosen to be broad and relatively uninformative.

Inter-individual differences in behavior are a main source of variance in eye movement data. Here we took advantage of these differences by testing model generalizability. We implemented individual independent model fitting for each participant by running a DREAM parameter estimation for each participant separately. The advantage of using this method is that when simulating data, we obtain an upper limit for the variance of parameters between individual participants.

2.5 Data availability

The experimental data used in this study represent a subset of the Potsdam Corpus on Spatial Frequency Search in Natural Scenes (Rothkegel, Schütt, et al., 2019), which is publicly available via the Open Science Framework (`osf.io/caqt2`).

2.6 Code availability

Source code used for numerical simulations, analyses, and plotting as well as other project-related files are made available (`osf.io/qsx4w`).

2.7 Acknowledgements

We thank Noa Malem-Shinitski, Maximilian Rabe, Stefan A. Seelig, Silvia Makowski for valuable discussions. This work was supported by a grant from Deutsche Forschungsgemeinschaft, Germany (SFB 1294, project no. 318763901). We acknowledge a grant for computing time from Norddeutscher Verbund für Hoch- und Höchstleistungsrechnen (HLRN), Germany (grant no. bbx00001).

3 Modeling Task influences

A dynamical scan path model for task-dependence during scene viewing

Lisa Schwetlick, Daniel Backhaus, Ralf Engbert
University of Potsdam

Abstract

In real-world scene perception human observers generate sequences of fixations to move image patches into the high-acuity center of the visual field. Models of visual attention developed over the last 25 years aim to predict two-dimensional probabilities of gaze positions for a given image via saliency maps. Recently, progress has been made on models for the generation of scan paths under the constraints of saliency as well as attentional and oculomotor restrictions. Experimental research demonstrated that task constraints can have a strong impact on viewing behavior. Here we propose a scan path model for both fixation positions and fixation durations, which includes influences of task instructions and interindividual differences. Based on an eye-movement experiment with four different task conditions, we estimated model parameters for each individual observer and task condition using a fully Bayesian dynamical modeling framework using a joint spatial-temporal likelihood approach with sequential estimation. Resulting parameter values demonstrate that model properties such as the attentional span are adjusted to task requirements. Posterior predictive checks indicate that our dynamical model can reproduce task differences in scan path statistics across individual observers.

Published in **Psychological Review, 2022**

3.1 Introduction

From the early days of eye movement research into the present, the question of how task influences the decisions on and the order of fixation locations has been of central interest. One of the first eye movement studies, the seminal but anecdotal work by Yarbus (1967) suggests qualitative differences in scan paths when looking at the same image under different task instructions. Yarbus concluded that both the fixation density and sequences of fixated locations (i.e., scan paths) sensitively depend on task requirements. Within the large body of subsequent work (see below) on this topic, a variety of methods for investigating the role of task for active vision (Findlay & Gilchrist, 2003) have been proposed. Comparisons of eye movement measures demonstrate that spatial fixation locations as well as fixation durations are influenced by task (Castelhano et al., 2009). It was also noted that differences between tasks can be larger than the interindividual differences between observers (DeAngelus & Pelz, 2009).

In this paper we study a theoretical model to investigate the research question of how task demands modulate scan path generation. Modeling scan path generation provides crucial constraints on underlying cognitive, attentional, and motor processes (e.g., Engbert, Trukenbrod, et al., 2015; Le Meur & Liu, 2015; Schwetlick, Rothkegel, Trukenbrod, et al., 2020b; Tatler et al., 2017). Here we develop and analyze a mathematical model of scan path generation across tasks. First, we advance our earlier dynamical model (Engbert, Trukenbrod, et al., 2015; Schwetlick, Rothkegel, Trukenbrod, et al., 2020b) to include the control of fixation durations in addition to fixation positions. Second, the model will be fitted to experimental scan paths from individual observers using Bayesian inference for dynamical models (Engbert et al., 2022; Schütt et al., 2017). Third, with this detailed account of scan path generation, we model task-dependence across four different viewing conditions from an earlier experimental study (Backhaus et al., 2020).

3.1.1 Task differences in scene viewing

Early reports by Buswell (1935) and Yarbus (1967) lend support to the idea that eye movement patterns depend on the viewer's instruction and not just on image content and features. Such effects of top-down task impact on viewing behavior have been replicated by follow-up experimental studies (Castelhano et al., 2009; DeAngelus & Pelz, 2009; Mills et al., 2011). Related work included investigations of eye movements during everyday tasks like preparing a cup of tea (Land et al., 1999) or a sandwich (Hayhoe et al., 2003). During such tasks, gaze control supports general motor control by either moving relevant information to the central visual field (Ballard et al., 1997; Land & Tatler, 2009) or by selecting object information needed later during the task to prepare future movements (Pelz & Canosa, 2001). Important examples include driving (Land & Hayhoe, 2001), cycling (Vansteenkiste et al., 2014), walking (Matthis et al., 2018; Rothkopf et al., 2007), and ball games (Land & Furneaux, 1997; Land & McLeod, 2000). These experimental designs move away from the typical lab-based scene viewing

paradigm and contribute to a more ecologically valid account of eye guidance. In the typical scene viewing paradigm where no task is given, participants are free to choose their own objective or task, which is hidden from the researcher's access (Tatler et al., 2011). Given the relevance of specific viewing strategies to different tasks, scene viewing without clear task instruction might thus be difficult to interpret.

The rise of modern machine learning techniques motivated purely data-driven research on scan path patterns to identify task from experimental fixation sequences. Initially, work on this topic generated mixed results (see Boisvert & Bruce, 2016, for a detailed review). Based on scan path visualization of their underlying data, Greene et al. (2012) found that neither human experts nor any of three proposed pattern classifiers were able to reliably infer which task the observer was performing. The same experimental data were later reanalyzed by Borji and Itti (2014). As a result, the classifier could be improved significantly, showing 35% accuracy for a four-task classification data set, where the reanalysis included more spatial data in the form of low resolution fixation density patterns. Performance was further boosted by accounting for inter-individual and image differences (Kanan et al., 2014). Furthermore, a classifier trained using a hidden Markov model approach indicated that additional diagnostic information for successful task prediction is contained in the scan path dynamics (Haji-Abolhassani & Clark, 2014).

Experimental results agree that the given task significantly affects gaze characteristics. Specifically, Castelhano et al. (2009) found that both number of fixations and fixation durations varied with task and that fixated areas were qualitatively different between tasks. Other studies also found effects of task on temporal (e.g., fixation duration) as well as spatial (e.g., saccade amplitude) measures (Bonev et al., 2013; Mills et al., 2011). Search tasks have also been found to lead to an extended range of fixation locations compared to free viewing material (Tatler, 2007). However, finding systematic differences for the type of task, such as free viewing or search has yielded inconsistent results. While Mills et al. (2011) found shorter fixation durations for search tasks compared to free viewing, results disagree about saccade amplitude with more recent findings by Backhaus et al. (2020). Because of the variety of tasks and stimuli in experimental paradigms, however, it can be expected that comparisons of results across studies is not straightforward and do not always lead to full agreement.

Taken together, experimental work as well as machine-learning classification paint a consistent picture that individual differences, spatial selection, and aggregate eye movement measures are specific for tasks. Classification success depends highly on the particular set of features selected and on the type of classification algorithm used. A number of studies used features of varying abstraction, from the very basic saccade amplitude and fixation duration to global or local image features (Boisvert & Bruce, 2016) or transition probabilities between identified regions of interest (Coutrot et al., 2017). In the next sections we discuss the role of process-oriented models in scene-viewing research, with a particular focus on scan path generation.

3.1.2 Theoretical models of visual attention during scene viewing

Human eye movements in natural scene viewing are guided by visual attention (Itti & Koch, 2001), which is modulated by image-dependent features. Basic research showed that saccadic eye movements follow the locus of attention (Deubel & Schneider, 1996; Kowler & Blaser, 1995). This tight coupling of attention and saccades is exploited in experimental work, where gaze positions are typically equated with the locus of visual attention (Henderson, 2003). It should be noted, however, that there are pronounced deviations between visual attention and gaze position around the time of saccade (Deubel & Schneider, 1996; Kowler & Blaser, 1995). For example, we recently showed in a mathematical modeling study that effects of perisaccadic attention can explain effects of saccade statistics in scene viewing (Schwetlick, Rothkegel, Trukenbrod, et al., 2020b) within the SceneWalk model (Engbert, Trukenbrod, et al., 2015, see below). The following section discusses primarily image-computable models for the spatial, time-averaged distribution of fixation positions as a proxy for visual attention, which constitute a large part of the literature on the topic.

Past modeling work shows that image-dependent influences contribute strongly to predictions of the overall gaze positions when viewing natural scenes. Relevant image features include local luminance contrast and edge density (Mannan et al., 1997; Parkhurst & Niebur, 2003; Reinagel & Zador, 1999; Tatler et al., 2005). Much of the early modeling work to predict fixations involved the detection of these features in images to generate an activation map, in which high activation represents areas rich in features relevant for the viewing task.

In 1985, Koch and Ullman proposed a model that (i) simultaneously computes various feature maps for color, intensity, edges, and feature popout detected by biologically plausible components such as center-surround receptive fields and that (ii) integrates these maps into a master map called *saliency map*. This idea was later implemented as a computational model (see Itti & Koch, 2001, for an overview). More feature-based saliency models followed, adding other spatial features or statistics (e.g., Bruce and Tsotsos, 2009; Harel et al., 2006; Tatler et al., 2006; Torralba et al., 2006).

One way of evaluating saliency models is to compare the generated saliency maps to empirical fixation densities from eye movement experiments (Parkhurst et al., 2002; Tatler et al., 2005). Linking the two concepts is the assumption that attention and eye movement are closely related (Henderson, 2003), as discussed above. In order to be able to compare a two-dimensional density and a series of fixation locations, metrics of different sophistication have been proposed (Kümmerer et al., 2015), including AUC score (Tatler et al., 2005) and Kulback-Leibler divergence (Le Meur & Liu, 2015). The popular MIT/Tübingen Saliency Benchmark[4] employs seven such metrics, each emphasizing different aspects and yielding different results. More recently, Kümmerer et al. (2015) suggested to use the concept of information gain based on statistical model likelihood applied to saliency modeling, as a statistically well-founded alternative to ad-hoc metrics.

4 See https://saliency.tuebingen.ai

Beyond simple feature-based attention, processes of target selection can also include higher level concepts such as objects or contextual guidance (Nuthmann & Henderson, 2010; Torralba et al., 2006). A long-standing debate exists between the more top-down interpretation that emphasizes cognitive relevance and meaning (Henderson et al., 2007; Henderson & Hayes, 2017; Henderson et al., 2019; Henderson & Smith, 2009) and the traditional saliency oriented approach (Pedziwiatr et al., 2021a). The distinction is not as clear-cut as some discussions suggest, however, as clusters of low-level image features are also indicative of objects. Research does show that incorporating information about object locations into saliency models makes them more accurate (Kümmerer, Theis, et al., 2014). Thus, eye movements are driven both by basic features detected by low-level vision as well as more advanced levels of cognitive processing (Schütt et al., 2019).

Using deep learning techniques and neural network architectures, recent models of general visual saliency achieved considerably improved prediction accuracy (Bylinskii et al., 2015; Kümmerer et al., 2017) compared to earlier approaches. Neural network models are trained using experimental fixation data, which represent image-driven as well as meaning- and task-dependent influences. Correspondingly, although the outcomes are referred to as saliency maps, models trained using this data-driven approach produce descriptions of eye movements which can no longer be dissociated into particular layers of cognition.

It is also worth noting that almost all experimental data used in this data-driven approach are acquired from participants who viewed pictures without a specific task. The underlying assumption is that simple picture viewing results in the most natural behavior. This was criticized by Tatler et al. (2011): "It seems more likely that free viewing tasks simply give the subject free license to select his or her own internal agendas" (p. 4).

Besides the viewing task, another aspect of ecologically valid, real-world conditions is the possibility of body movement (Backhaus et al., 2020; Matthis et al., 2018). In order to evaluate whether typical lab restrictions, e.g., head stabilization on a chin rest, limit the generalizability of results (Tatler et al., 2011). Backhaus et al. (2020) suggested the use of mobile eye-tracking which permitted natural posture and postural fluctuations (Collins & De Luca, 1995), with the general finding that task effects are robust with respect to changes in body posture while viewing images.

Modeling of visual attention during scene viewing often focuses exclusively on the spatial aspect of *where* fixations are placed on an image. Next, we discuss how time-ordered fixation sequences are generated and which factors influence scan path dynamics via biologically inspired mechanisms—including the timing of saccades (Henderson, 2003).

3.1.3 Biologically inspired models of scan path generation

During active vision (Findlay & Gilchrist, 2003), our gaze continually explores the visual environment by producing saccadic movements. Findlay and Walker (1999a) proposed an influential conceptual model for generating saccades, which claims validity

for a variety of experimental paradigms and situations. A fundamental assumption in the model is that two partially separate pathways exist for temporal and spatial control processes. Both pathways are composed of a hierarchy of levels from automatic to higher-level cognitive control, where each level has high biological plausibility. This basic architecture turned out to be successful in a variety of cognitive tasks, such as reading (Engbert et al., 2002; Engbert et al., 2005; Rabe et al., 2021; Seelig et al., 2020) and scene viewing (Engbert, Trukenbrod, et al., 2015; Schwetlick, Rothkegel, Trukenbrod, et al., 2020b). Thus, there is theoretical support for the existence of separate pathways for to spatial and temporal control of gaze.

As discussed in the previous section, significant progress has been made on models that predict the spatial control of gaze position, often termed visual saliency modeling (Itti & Koch, 2001). These models aim at predicting the 2D density of fixations on a given scene. From the beginnings of the research tradition, biological plausibility played an important role for the development of these models (Koch & Ullman, 1985). However, initially there was little interest in making use of saccade statistics that were more detailed than the spatial density of fixations.

While the role of scan paths (Yarbus, 1967) and sequential effects in sequences of saccades (Noton & Stark, 1971a, 1971b) was noted early on, it was only much later that the dissatisfaction with purely saliency-based models stimulated interest in scan path generation (Zelinsky, 2008). For example, Tatler and Vincent (2009) demonstrated that adding oculomotor principles to models could significantly improve the predictive power in spatial selection. Thus, effects of the previous fixation location on the selection of the upcoming gaze position were identified as an important modeling goal (Le Meur & Liu, 2015) for understanding principles of human gaze control.

The success of mathematical models to reproduce human gaze positions stimulated interest in theoretical models for fixation durations (Henderson, 2003). Since each fixation is bounded by two saccades, effects of oculomotor preparation and execution are highly relevant to the statistics of fixation durations and saccade timing. Nuthmann and Henderson (2010) proposed a random-walk model for timing of saccades (CRISP) to explain data from an experimental paradigm with delayed scene onsets. Similarly, Tatler et al. (2017) used a activation-based rise-to-threshold unit (Reddi & Carpenter, 2000) for the generation of saccadic onset-times in their model (LATEST). The LATEST model (Tatler et al., 2017) is a combined model of spatial and temporal control. This approach represents of mixture of process-oriented (LATER unit; Reddi & Carpenter, 2000) and data-driven modeling (spatial aspects). As a result, the LATEST model demonstrates important effects on the integration of spatial and temporal control of saccades. However, the data-driven components of LATEST offer limited insight into the biological processes that generate the behavior.

A recent paper that contributed to the conceptual advancement in the field of eye-movement modeling was published by Kucharsky et al. (2021). In this theoretical study, a model for fixation durations from the broad class of information accumulation or drift-diffusion models (P. L. Smith & Ratcliff, 2004) was extended by a spatial component. The integrated model termed WALD-EM was successful in modeling many aspects of saccade statistics and distributions of fixation durations. Most im-

portant for theory building, a combined spatiotemporal likelihood function was used, which provides the basis for rigorous statistical inference.

Statistical, functional, and mechanistic modeling in cognitive science take on vastly different roles (Bechtel & Abrahamsen, 2010). While statistical models are mainly descriptive, functional and mechanistic models are process-oriented and propose specific interactions between different subsystems. Thus, specific assumptions can be tested against experimental data, so that the plausibility of biologically-inspired mechanisms can be tested (Engbert, 2021). The more grounded in experimental evidence and the more mechanistic the model, the more compelling are the conclusions are about the explanation of an observed effect (Bechtel & Abrahamsen, 2010). Moreover, in mechanistic, generative models it is possible to interpret the model parameters with respect to the processes in the visual, attentional, and oculomotor systems. Within this class of models we developed the SceneWalk model (Engbert, Trukenbrod, et al., 2015; Schütt et al., 2017; Schwetlick, Rothkegel, Trukenbrod, et al., 2020b), which is in agreement with the framework proposed by Findlay and Walker (1999a).

In the SceneWalk model, fixation selection is based on a time-dependent priority map that is influenced by the current gaze position, time-independent fixation density, and previously fixated locations. More recently, Schwetlick, Rothkegel, Trukenbrod, et al. (2020b) added perisaccadic attentional processes that improved the model's performance with respect to a variety of scan path metrics including complex effects such as modulations of the mean fixation durations by saccade turning angle. With respect to task influences discussed in the current work, we expect that differences in scan paths across tasks will be reflected in differences of the numerical values of model parameters across tasks. The ability of dynamical process-oriented models to reproduce differences in behavior via parameter adaptation supports the underlying mechanisms by demonstrating generalizablity. Our assumptions of how parameter values vary between tasks is based on prior experimental work showing that repeated viewing of the same natural scenes induce differences in saccade statistics (e.g., distribution of saccade lengths). These differences are compatible, in the model, with a smaller perceptual span during second viewing of the same image compared to the first viewing (Trukenbrod et al., 2019).

3.1.4 The role of saliency for dynamics

Due to intense research, the scientific literature on modeling of static saliency (2D fixation density) has grown enormously, while scan path modeling is a comparatively new field of quantitative modeling. In this paper we use the SceneWalk model of eye movement dynamics (Engbert, Trukenbrod, et al., 2015; Schwetlick, Rothkegel, Trukenbrod, et al., 2020b) to investigate the principles of task-dependent scan paths. In order to generate scan paths, the model relies on activation maps which approximate visual saliency. As a stable upper bound, in earlier studies we used experimental fixation densities. Alternatively, the model could also be combined with a saliency model and generate eye movements from computer-generated saliency maps.

The aim of the current study is to investigate the modulation of underlying processes

of static and dynamic components of eye guidance caused by task variation. It is important to note that the term *saliency map* has been used to describe different concepts in the literature. As discussed above, visual saliency initially referred to very low-level image features like edges (Itti & Koch, 2001). This early concept of saliency is completely devoid of higher level influences such as task. Later, however, the concept of saliency was expanded to include all influences that improve predictions on visual attention as indicated by gaze (Kümmerer, Theis, et al., 2014). It is clear that if complex models are fitted to empirical fixation densities, then higher-level factors such as task are difficult to separate from low-level vision. Recently, there has been much discussion about the importance of high level image features versus meaning as a main predictor of eye movements (Henderson & Hayes, 2017, 2018; Henderson et al., 2019; Pedziwiatr et al., 2021a, 2021b). While this discussion is clearly relevant to the distinction between top-down and bottom-up influences in vision (Schütt et al., 2019), the focus of the current work is on the interaction between saliency, task, and the dynamics of scan path generation.

As discussed above there is empirical evidence that viewing strategies during scene viewing depend on the given task. One possibility is that the main cause for this difference is an adjustment of the prioritization of visual information. Based on this assumption, elements in the visual display are weighted by attention according to their importance to the task. This hypothesis requires the input saliency for the eye movement model to be separate *task-specific saliency* maps for each image and task. As an alternative hypothesis we might consider differences between tasks to be attributed to the tuning of saccade dynamics to particular tasks. Here, task-specific weighting of image features can be neglected and the eye movement model uses the same *general purpose saliency* input per image for all tasks. To represent this idea, we use a general fixation density from a free viewing task as a basis for task-specific model parameter estimation and validation. It seems likely that task-specific saliency effects as well as task-specific eye movement effects will play a role.

Here we present the results for two alternative models using general and using task-specific fixation densities as input to the eye movement model. We then compare the resulting performance of both models. The general fixation density in the model is related to the task-independent interpretation of saliency. The saliency map that is passed into the model is constructed using experimentally recorded scan paths from a separate free-viewing experiment using the same images (Backhaus & Engbert, 2022b). The task-specific saliency version of the model is identical, with the difference that separate fixation densities were constructed using gaze data from only one of four different task conditions.

With concurrent models for the same experimental data, model inference has become an increasingly important topic in cognitive modeling. Recently, the numerical tools and the computation power have become available to carry out rigorous parameter inference and model comparisons (Schütt et al., 2017; Schwetlick, Rothkegel, Trukenbrod, et al., 2020b), in particular, if the likelihood function for the model can be computed or approximated. In the next section, we discuss statistical inference for dynamical models.

3.1.5 Bayesian parameter inference for dynamical models

Dynamical models of eye-movement control generate specific predictions for sequential dependencies of fixations over time. As a consequence, the full potential of statistical inference for dynamical models unfolds if model predictions are evaluated based on fixation sequences (Engbert et al., 2022). This approach requires sequential predictions for upcoming fixations, advanced computational methods, and sufficient computing time (Schütt et al., 2017). Many state-of-the-art methods for parameter inference in cognitive models, however, are based on ad-hoc performance metrics (Engbert, Trukenbrod, et al., 2015; Le Meur & Liu, 2015; Tatler et al., 2017; Zhou & Yu, 2021). Often (but not always), such ad-hoc metrics ignore the sequential structure of scan paths. In these cases, researchers choose relevant metrics and compute a loss function that indicates how closely simulated data resemble experimental data based on the pre-defined metrics. Model parameters are obtained by optimization of the loss function when model parameters are varied. As a result, model inference is subjective (i.e., dependent on the choice of the loss function) and difficult to generalize, since arbitrary metrics will optimize the model to reproduce some aspects of the model while ignoring others.

A statistically well-founded alternative is based on the likelihood function $L_M(\theta|\text{data})$ of a model M with parameters θ given an experimental data set (Myung, 2003). The likelihood is defined as the conditional probability P_M for observing the data in the context of model M specified by parameters θ, i.e.,

$$L_M(\theta|\mathrm{d}ata) = P_M(\mathrm{d}ata|\theta) \; . \tag{3.1}$$

If numerical computation of the likelihood, Eq. (3.1) is possible, then rigorous statistical inference on model parameters and comparisons between different models are also possible, including Bayesian inference (Gelman et al., 2013). In static saliency modeling, the use of the likelihood (Kümmerer et al., 2015) is straightforward: the saliency map is interpreted as a fixation probability and the probability of each experimentally observed fixation position is evaluated on this probability map. In dynamical scan path modeling the process is more elaborate as explained in the following (Engbert et al., 2022; Schütt et al., 2017).

A fixation f_i in a scan path $\mathcal{F} = \{f_1, f_2, f_3, ..., f_N\}$ is given by its spatial position (x_i, y_i) and its fixation duration T_i. Thus, a fixation $f_i = (x_i, y_i, T_i)$ is a 3-tuple. Because of the sequential nature of the scan path, the likelihood can be decomposed into a product of conditional probabilities, i.e.,

$$\begin{aligned} L_M(\theta|\mathcal{F}) &= L_M(\theta|f_1, f_2, f_3, ..., f_N) & (3.2)\\ &= P_M(f_1|\theta) \prod_{i=2}^{N} P_M(f_i|f_1, f_2, ..., f_{i-1}; \theta) \; , & (3.3) \end{aligned}$$

where the generative model is used to estimate the probability $P_M(f_i|f_1, f_2, ..., f_{i-1}; \theta)$ of the ith fixation when enforcing the previous fixations $f_1, f_2, ..., f_{i-1}$ and $P_M(f_1|\theta)$ is

the first fixation that is typically known and experimentally controlled (Schütt et al., 2017; Seelig et al., 2020), so that $P_M(f_1|\theta) = 1$.

In Bayesian inference, we specify a prior probability $P(\theta)$ over the model parameters and use the likelihood $L_M(\theta|\mathcal{F})$ to compute the posterior probability $P(\theta|\mathcal{F})$ using Bayes' theorem,

$$P(\theta|\mathcal{F}) = \frac{L_M(\theta|\mathcal{F})P(\theta)}{\int_\Omega L_M(\theta|\mathcal{F})P(\theta)\mathrm{d}\theta} \,. \tag{3.4}$$

The integral in the denominator in Eq. (3.4) is typically intractable for realistic cognitive models. Therefore, the posterior probability $P(\theta|\mathcal{F})$ is estimated numerically via Markov Chain Monte Carlo methods (Gilks et al., 1996). We will discuss the specific numerical procedures for parameter estimation in the methods sections.

3.1.6 The current study

The research goal of this study was to carry out model-based analyses of task effects on scan path generation. The starting point for our modeling work will be the SceneWalk model for scan path generation during scene viewing (Engbert, Trukenbrod, et al., 2015; Schütt et al., 2017; Schwetlick, Rothkegel, Trukenbrod, et al., 2020b). Recently, we included peri-saccadic attentional effects, which reproduced correlations between saccade turning angles with saccade lengths and fixation durations (Schwetlick, Rothkegel, Trukenbrod, et al., 2020b). This variation of the SceneWalk model can reproduce systematic variations in mean fixation durations. However, the peri-saccadic principles do not represent an explicit timing mechanism for saccades as proposed by Nuthmann and Henderson (2010) (see also Laubrock et al., 2013; Tatler et al., 2017).

Since explicit timing effects can be expected in task-dependent scene viewing, here we developed a further version of the model that includes a mechanistic timer. As shown by the LATEST model (Tatler et al., 2017), the saliency value at fixation exerts a negative effect on the decision rate, which translates into prolonged mean fixation durations for fixation location with higher saliencies compared to lower saliences. In order to investigate a coupling between temporal and spatial information in the SceneWalk model (Schwetlick, Rothkegel, Trukenbrod, et al., 2020b), we introduce a timing mechanism that enables the local saliency to influence mean fixation durations. This addition combines fixation durations and fixation locations into one coherent model, improving our general framework for generating fixation sequences. Coupling parameters for the spatial and temporal components are estimated from experimental data.

The structure of the manuscript is as follows. We start with a detailed explanation of the SceneWalk model and its underlying activation dynamics with activation and inhibition pathways (Schwetlick, Rothkegel, Trukenbrod, et al., 2020b). Next, we extend the model to include the explicit timing mechanism for saccade generation. The likelihood function of the extended model can be decomposed into a spatial and a temporal component. We discuss the specific approach for numerical Bayesian

inference. After the introduction to the model, we describe the experiment on natural scene-viewing which included a task manipulation (Backhaus et al., 2020). The Results show parameter estimation and posterior predictive checks that indicate the goodness-of-fit for various dependent variables. Based on the posterior estimates of the model parameters, we run a statistical analysis across participants and tasks that highlight how the model explained task effects. For the different model variations, we present results from likelihood-based model comparisons. A statistical analysis is also applied to generated scan path data and compared to experimental data, which confirm adequacy of the model fits. Finally, we discuss our results with respect to task dependence on scene-viewing and saliency modeling, inter-individual differences, and more general aspects on process-oriented modeling.

3.2 SceneWalk: A framework for dynamical scan-path modeling

The SceneWalk model (Engbert, Trukenbrod, et al., 2015; Schütt et al., 2017) implements two largely independent processing streams: one activatory and one inhibitory (see Figure 3.1). Both streams are grounded in theoretical (Itti & Koch, 2001) and experimental work (Rothkegel et al., 2016), showing that they represent the two main factors contributing to fixation selection.

The activation stream combines information about image features with a mechanism for foveation and thereby yields an approximation of the information that can be extracted from an image at a particular fixation location. Image features include edge and contrast information as well as more high level information such as objects. This information is passed to the SceneWalk model in the form of a normalized fixation probability map. The SceneWalk model is solely a model of dynamics and requires a saliency map to be provided as input for each image. This map could be computed by one of the implemented saliency models (e.g. Kümmerer et al., 2015) that follow the general modeling approach (Itti & Koch, 2001). For the later interpretation of results, strengths and weaknesses of the saliency models must be separated from shortcomings of the scan path model. Therefore, we will use the experimentally observed fixation density estimate, which represents the theoretical upper limit for the performance of the salience model. Mismatches between data and model output are therefore predominantly caused by the scan path model, although the time-averaging assumption for the fixation density is another approximation that may contribute. The second component of the activation stream is related to the visually attended region. When attention is aligned with the current fixation position, the decreasing receptor density of the retina towards the periphery leads to a decline in visual acuity, which we implement as a Gaussian window centered around the current fixation. For attention shifts to the periphery, discussed below, we keep the Gaussian window approximation to implement an attentional spotlight on an upcoming target (Engbert et al., 2011; Itti & Koch, 2001; Shulman et al., 1979; Tsotsos, 1990). The convolution of saliency and the

Gaussian window results in the input to the activation stream (Engbert, Trukenbrod, et al., 2015).

The inhibition stream of the SceneWalk model is responsible for fixation tagging, i.e., keeping track of fixated regions and preventing the continuous return to the same high saliency regions (Bays & Husain, 2012; Klein, 2000). Evidence from visual search (Posner et al., 1985) and also scene viewing (Bays & Husain, 2012; Klein & MacInnes, 1999; Rothkegel et al., 2016) and electrophysiology (Hopfinger & Mangun, 1998; Mirpour et al., 2019) shows the relevance of inhibition of return for scan path statistics. The input to the inhibition stream for fixation tagging is also implemented as a Gaussian centered at the current fixation location (see Figure 3.1).

The two separate streams evolve continuously over time. Among the previous modeling results, we found that the build-up of activation in the inhibition stream is slower than the activation stream (Engbert, Trukenbrod, et al., 2015). As a consequence, inhibitory tagging evolves slowly, so that refixations of recently fixated scene regions are still possible (T. J. Smith & Henderson, 2009). In the activation stream of the most recent version of the model (Schwetlick, Rothkegel, Trukenbrod, et al., 2020b), we added a directed, smaller facilitation of return (Luke et al., 2013; T. J. Smith & Henderson, 2009) in addition to the slow, global inhibition of return. The interplay of both mechanisms results in a slower decay of activation at the previous location and briefly enables precisely directed return saccades. Thus inhibition of return, attention, and facilitation of return can coexist by separating their temporal dependence. The extended model is described in more detail in the following section. The combination of the activation and inhibitory streams yields a priority map for saccade targeting (Bisley & Mirpour, 2019), which the model uses as the 2D fixation probability map for the selection of the upcoming saccade target. In the following, we discuss the dynamical behavior of both activation and inhibition streams as well as its combination to generate a priority map, which depends on fixation history and indicates the time-dependent probability of the target selection process.

3.2.1 Activation dynamics of attention and inhibitory fixation tagging

The most recent version of the SceneWalk model (Schwetlick, Rothkegel, Trukenbrod, et al., 2020b) implements the two-stream architecture discussed above as well as perisaccadic attentional mechanisms, which are related to saccade preparation and execution. As in the original model (Engbert, Trukenbrod, et al., 2015; Schütt et al., 2017), the activation and inhibition streams evolve over time and are combined mathematically to yield a moment-to-moment priority map (Bisley & Mirpour, 2019), from which target locations are selected probabilistically.

The model is implemented on a 128×128 grid, where (x, y) give the physical coordinates in degrees of visual angle. The inhibition/fixational tagging pathway is defined as a 2-D Gaussian centered around the current fixation position (f_x, f_y). It evolves

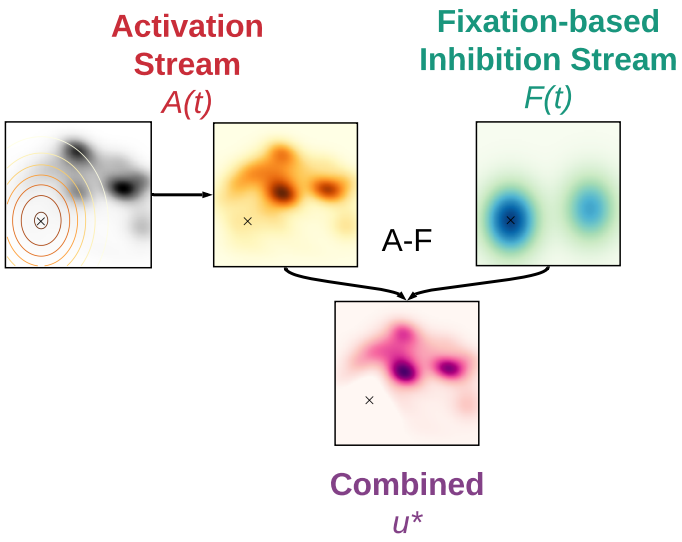

Figure 3.1 Two-stream architecture of visual attention and inhibitory tagging. The current fixation position is marked by the symbol "x". The streams evolve neural activations independently over time depending on the fixation position, input and decay. The activation stream receives as input a saliency map (black and white color map) which is convolved with a Gaussian aperture to approximate the visual attention span (orange color maps). The blue color maps represent inhibitory fixation tagging, which keeps track of previously visited locations. When both maps are combined the result is a priority map we interpret as the fixation selection probability.

over the duration of the fixation according to the differential equation

$$\frac{\mathrm{d}F_{ij}(t)}{\mathrm{d}t} = \omega_F \left(\frac{G_F(x_i, y_j; x_f, y_f)}{\sum_{kl} G_F(x_k, y_l; x_f, y_f)} - F_{ij}(t) \right) , \tag{3.5}$$

where F, denotes the fixation-based inhibition stream, G_F is the Gaussian-shaped activation window with standard deviation parameter σ_F, and ω_F is the parameter for the speed of decay.

The activation stream is implemented as a separate ODE, with its own separate time scale. It similarly includes a Gaussian window around the fixation location, emulating the decrease of visual acuity towards the periphery, and includes information about visual saliency, which must be passed to the model. The differential equation for the activation stream is given by

$$\frac{\mathrm{d}A_{ij}(t)}{\mathrm{d}t} = \omega_A \left(\frac{S_{ij}\, G_A(x_i, y_j; x_f, y_f)}{\sum_{kl} S_{kl}\, G_A(x_k, y_l; x_f, y_f)} - A_{ij}(t) \right) , \tag{3.6}$$

where A, denotes the activation stream, G_A is the Gaussian-shaped activation window with size σ_A, centered around the appropriate location for each phase, S is the saliency map of the image, and ω_A is the parameter for the speed of decay. The computation of numerical solutions of the of ODEs given by Eqs. (3.5-3.6) for all grid points (i, j)

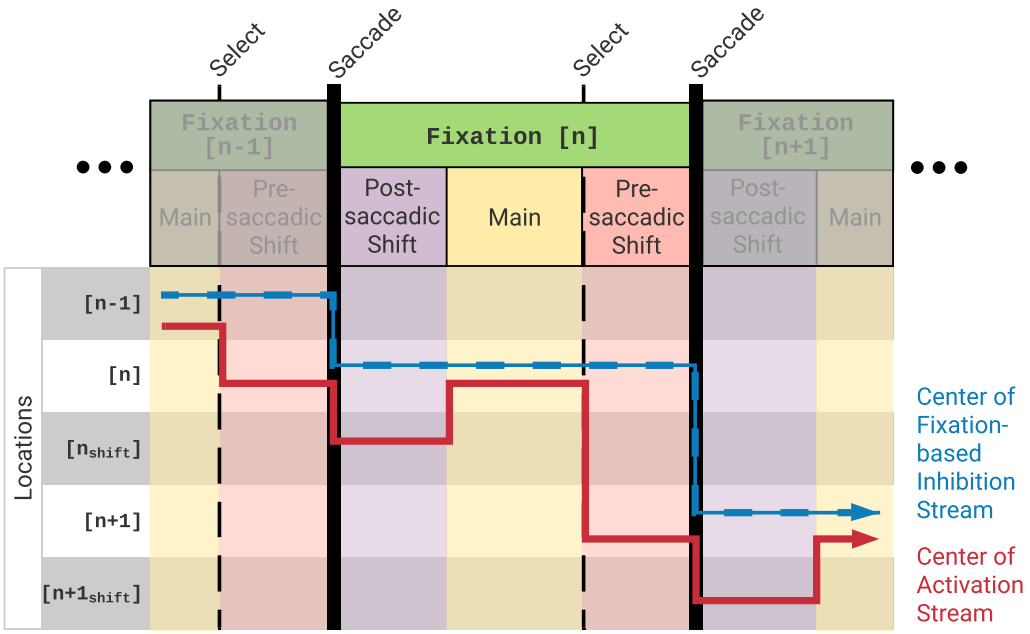

Figure 3.2 Temporal sequence of the three phases peri-saccadic attention. In the SceneWalk model, each fixation is split into three phases. During the main phase (yellow color) attention and fixation location are aligned, so that the activatory (red, solid line) and the inhibitory (blue, dashed line) Gaussian inputs are both centered around the current fixation position. The main phase is followed by a pre-saccadic shift (rose color) , where the attention precedes the eye position to the selected location. After each saccade (black line) a brief post-saccadic shift (purple color) causes the attention to be shifted further along the saccade vector, before fixation position and attention once again align at the new fixation position.

is discussed in Appendix B.2.

The extended model version (Schwetlick, Rothkegel, Trukenbrod, et al., 2020b) adds changes around the time of saccade to the model, where the temporal aspect is illustrated in Figure 3.2. Each fixation is split into three distinct phases: main phase, pre-saccadic shift, and post-saccadic shift. The rationale behind the extension is that before each saccade, attention precedes the eye to the target location. After a saccade has been executed, research shows evidence for a brief shift to account for the post-saccadic retinotopic attention trace. Thus, in the extended model, the center of the attentional Gaussian does not always align with the fixation position (overt attention), but instead the two decouple around the time of saccade. Previous work has shown that these components of saccade generation improve important statistical properties of the predicted scan path (Schwetlick, Rothkegel, Trukenbrod, et al., 2020b).

The saccade direction is given by the vector (x_δ, y_δ) with $x_\delta = x_n - x_{n-1}$ and $y_\delta = y_n - y_{n-1}$. Therefore, the location of the post-saccadic shift is determined by

$$(x_\mathrm{s}, y_\mathrm{s}) = (x_n, y_n) + \frac{(x_\delta, y_\delta)}{\sqrt{x_\delta^2 + y_\delta^2}} \cdot \eta \ , \tag{3.7}$$

i.e., the target region of the shift corresponds to a point along the previous saccade vector, where η determines the shift amplitude relative to the previous saccade length.

The peri-saccadic extension of the model requires that the differential equations for the evolution of activations are applied to the three phases for each saccade, since the center of the activation stream is in a different position at each step (see Figure 3.2).

In order to select the next fixation target a priority map is computed, by combining both streams. The exponent γ shapes this priority map, making it more deterministic, the higher the exponent becomes,

$$u_{ij}(t) = \frac{(A_{ij}(t))^\gamma}{\sum_{kl}(A_{kl}(t))^\gamma} - C_F \frac{(F_{ij}(t))^\gamma}{\sum_{kl}(F_{kl}(t))^\gamma} \,. \tag{3.8}$$

Negative values of u_{ij} indicate excess inhibitory activations, which render the saccade targeting probability zero. Thus, for computation of the saccade probability is based on the positive values u^*_{ij}, defined as

$$u^*_{ij} = \begin{cases} u_{ij} & \text{if} \quad u_{ij} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.9}$$

Finally, since zero fixation probability does not exist in real experiments, a noise term ζ is added to warrant fixation in regions with $u_{ij} < 0$ with low probability, i.e., the target selection probability at position (i, j) is given by

$$\pi(i,j) = (1-\zeta)\frac{u^*_{ij}}{\sum_{kl} u^*_{kl}} + \zeta \frac{1}{\sum_{kl} 1} \,. \tag{3.10}$$

The extended model also includes mechanisms for center bias and facilitation of return, for which we provide detailed mathematical equations in the Appendix.

3.2.2 Temporal control of fixation durations and coupling to local saliency

Because of the dynamical nature of the activation maps in the SceneWalk model, saccadic selection probabilities (or the priority map) change over time during fixations. Therefore, we clearly expect the model to predict interactions of temporal and spatial aspects of saccade preparation. This theoretical expectation is in good agreement with the results of statistical parameter fitting in the LATEST model (Tatler et al., 2017), which demonstrates various correlations between spatial and temporal aspects of saccade selection.

We assume that fixation durations are controlled by a continuous-time discrete-state random walk process (see also Laubrock et al., 2013; Nuthmann & Henderson, 2010). The distribution of fixation durations T generated by this random walk is a Gamma distribution, which can be written as

$$g(T) = \frac{b^q}{\Gamma(q)} T^{q-1} e^{-bT}, \tag{3.11}$$

with free parameters rate b and shape q. The mean fixation duration is given as $\mu_T = q/b$ and its variance is $\sigma_T^2 = q/b^2$. It is important to note that our model does not critically depend on the assumption of a Gamma distribution. The broad class of information accumulation or drift-diffusion models (P. L. Smith & Ratcliff, 2004) generates qualitatively very similar distributions. The WALD-EM model by Kucharsky et al. (2021) assumes a WALD (inverse Gaussian) distribution, which gives a comparable goodness-of-fit to the experimental data. Thus, the total duration of each fixation is generated by sampling from the Gamma distribution. According to the most recent version of the SceneWalk model (Schwetlick, Rothkegel, Trukenbrod, et al., 2020b), a fixation is subdivided into post-saccadic, main, and pre-saccadic phases. The full duration is therefore split into the three phases. The post- and pre-saccadic phases have fixed durations, which is an assumption inspired by experimental work on predictive allocation and remapping of attention (Rolfs et al., 2011). Specifically, the duration of the shifts were set to $\tau_{pre} = 0.05$ s for the pre-saccadic shift and $\tau_{post} = 0.1$ s for the post-saccadic shift, corresponding to the approximate durations found in the literature (Golomb et al., 2008; Rolfs et al., 2011).

An interesting and important question is if and how local saliency and mean fixation duration are related. Here we assume that mean fixation duration $\bar{T}_i$ at fixation location x_i parametrically depends on the logarithm of the local saliency $\log s(x_i)$. We assume that the shape parameter q of the distribution is constant, while the rate parameter b varies in relation to current input. Therefore, we will try to estimate the parameters t_α and t_β for a linear relationship between parameter b and the logarithm of the local saliency, i.e.,

$$b = t_\alpha + t_\beta \log s(x_i) \ . \tag{3.12}$$

In principle, we assume that the model's activation value at location x_i should be used in Eq. (3.12), not local saliency. For simplicity, the current version of our model uses the (logarithm of the) local saliency as an approximation for the average local activation.

In the following sections we will refer this, most recent, version of the model (with timing mechanism and attentional shifts) as SceneWalk. Previous versions of the SceneWalk model are not subject of this paper.

3.2.3 Full likelihood function for fixation positions and fixation durations

Previous versions of the SceneWalk model did not explicitly model saccade timing. With the gamma-distributed random-walk process for saccade triggering, we follow a strategy similar to the LATEST model (Tatler et al., 2017). In this section, we derive the full likelihood function of the model by including fixation durations in the likelihood function. A similar approach has been developed in the WALD-EM model (Kucharsky et al., 2021), who used a combined spatiotemporal likelihood. As discussed below, the spatial and temporal likelihood can be factorized, so that the log-likelihood sums up from spatial and temporal contributions.

A fixation i is determined by position x_i and fixation duration T_i, i.e., $f_i = (x_i, y_i, T_i)$. A scan path is a fixation sequence $\mathcal{F}_N = \{f_1, f_2, ..., f_N\}$ of N fixations. For an experimentally observed (or simulated) sequence of N fixations, the log-likelihood $l_M(\theta|\text{data})$ under model M specified by parameter vector θ is given by

$$l_M(\theta|\text{data}) = \sum_{i=1}^{N} \log P_M(f_i|\mathcal{F}_{i-1}, \theta) \ , \tag{3.13}$$

where $\mathcal{F}_{i-1}$ is the fixation sequence up to fixation $i-1$. The probability $P_M(f_i|\mathcal{F}_{i-1}, \theta)$ can be decomposed into a spatial (fixation location x_i) and temporal (fixation duration T_i) part, i.e.,

$$P(f_i|\mathcal{F}_{i-1}, \theta) = P^{spat}(x_i, y_i|\mathcal{F}_{i-1}, \theta) \cdot P^{temp}(T_i|x_i, y_i, \mathcal{F}_{i-1}, \theta) \ . \tag{3.14}$$

Therefore, the log-likelihood can be written as

$$l_M(\theta|\text{data}) = \sum_{i=1}^{N} \left(\log P_M^{spat}(x_i, y_i|\mathcal{F}_{i-1}, \theta) + \log P_M^{temp}(T_i|x_i, y_i, \mathcal{F}_{i-1}, \theta)\right) \ . \tag{3.15}$$

With the general procedure for sequential likelihood computation given by Eq. (3.2), we can write the log-likelihood of a full fixation sequence F_N as

$$l_M(\theta|\mathcal{F}_N) = \sum_{i=2}^{N} \log P_M(f_i|\mathcal{F}_{i-1;\theta}) \ . \tag{3.16}$$

Therefore, this spatio-temporal log-likelihood expands on and replaces the original, purely spatial likelihood function described in Schwetlick, Rothkegel, Trukenbrod, et al. (2020b).

3.2.4 Computational Bayesian inference of the SceneWalk model

With the computation of the model's likelihood function described in the previous section, Bayesian parameter inference can be implemented on a computer (Schütt et al., 2017). The advantage of the Bayesian framework is that we estimate not only point estimates for each parameter, but have access to the full posterior distribution over the model parameters. This is particularly desirable in models with a complex likelihood structure, where posteriors may be multi-modal or when we are interested in how much the data constrains the parameters (Gelman et al., 2013; Schad et al., 2021). Past studies have yielded promising results when applying Bayesian methods to dynamical cognitive models (e.g. Kucharsky et al., 2021; Rabe et al., 2021; Schütt et al., 2017; Seelig et al., 2020). For example, model parameters could be estimated for single participants, which was impossible before.

The most common numerical method for computation of the posterior is using Markov Chain Monte Carlo (MCMC) sampling (Gilks et al., 1996). A version of

this general approach is a random walk that samples higher density regions of the target distribution more frequently than lower density regions (Brooks et al., 2011). Beginning in a random location, the algorithm selects a candidate point according to a proposal distribution around the current location. This point can then be accepted or rejected based on the likelihood value at that location. It is important to note that even low probability points can be accepted. Thus, the algorithm proportionally samples the target distribution (Brooks et al., 2011).

In the present study, we applied the differential evolution adaptive Metropolis (DREAM) algorithm (Vrugt & Braak, 2011), which is a general-purpose MCMC sampler with excellent performance on complex, multimodal problems. The DREAM algorithm runs multiple Markov Chains in parallel, which can exchange information about past states. The latest version MT-DREAM(ZS) combines the strengths of multiple-try sampling, snooker updating, and sampling from an archive of past states (Laloy & Vrugt, 2012). These improvements help to optimize the convergence rate and also reduce the probability of individual chains running out of bounds or getting caught in local maxima. Recently, we applied the DREAM(ZS) algorithm successfully to the previous model version (Schwetlick, Rothkegel, Trukenbrod, et al., 2020b).

For the purposes of examining the differences between tasks, we split the data into a training and a test set. Thus, for each participant and task a randomized subset of 3/4 of the trials are considered training data and 1/4 is considered test data. For the parameter inference, we use training data. The sequential likelihood for each fixation in the training data is calculated for each point in the parameter space sampled by the estimation algorithm.

We estimate a subset of all model parameters that turned out to be critical for reproducing the most important statistics in experimental scan paths during previous studies (Schwetlick, Rothkegel, Trukenbrod, et al., 2020b). An overview of all fitted model parameters is given in Table 3.1. Priors were informed by the previous work with the model on other data sets. We used truncated Gaussian distributions as priors and kept them relatively uninformative in order to allow the data to constrain the model freely for each subject. The prior parameters are also reported in Table 3.1.

3.3 Experiment

With the theoretical extension of the SceneWalk model to generate fixation durations via explicit timing of saccades we set out to investigate a model-based explanation of task effects during scene viewing. The experimental data are taken from a recently published paper that report results from a paradigm with different viewing tasks (Backhaus et al., 2020). The experimental study includes eye-tracking data from 32 participants with normal or corrected-to-normal vision in a scene viewing experiment. Participants were asked to solve four different tasks while viewing 30 natural images.

Here, we focus on a basic description of the viewing tasks. For more details about the original experiment see Appendix Section B.1. Participants were required to count the

Parameter	Description	Eq.	Range	Mean	SD
ω_A	Speed of decay of the activation stream	(3.6)	0 ... 100	10	12
σ_A	Standard deviation of the Gaussian activation (°)	(3.6)	0 ... 30	7	5
σ_F	Standard deviation of the Gaussian inhibition (°)	(3.5)	0 ... 30	4	4
γ_i	Exponent regulating determinism in target selection	(3.8)	0 ... 5	1	3
$\log_{10}\zeta$	Noise parameter for target selection	(3.10)	−10 ... 0	−2	2
η	Size of the post-saccadic shift relative to saccade length	(3.7)	0 ... 4	0.5	2
t_α	Timing intercept	(3.12)	0 ... 5	3	5
t_β	Factor for the coupling of saliency and timer	(3.12)	−4 ... 0	−0.4	3
q	Shape parameter for the timing distribution	(3.11)	0 ... 15	3	3

Table 3.1 Model parameters for numerical inference. Range, mean, and standard deviation (SD) specify the truncated Gaussian priors for each parameter.

number of people in the scene images (Count People). Each image contained between 0 and 9 people; in some cases people were well hidden in the pictures. Another count task was to determine the number of animals shown in the image (Count Animals). Again, the number could vary between 0 and 9 animals. Since animals can appear in very different shapes and places compared to humans, the authors assumed that counting animals is the more difficult task. Both counting tasks share some characteristics of search tasks (Backhaus et al., 2020), because of the necessary detection of object type before counting.

The remaining two tasks investigated by Backhaus et al. (2020) are more unspecific with respect to the relevant scene regions, since in these tasks, participants were asked to guess the time of the day an image was taken (Guess Time) and to guess the country where the image was taken (Guess Country). The authors expected that light and illumination, the actions shown in the image (e.g., having lunch) but also clothing could give clues to the time of day or country of origin. These less specific viewing tasks might be looked upon as mildly constrained free-viewing tasks, while the more specific counting tasks might be considered as approximations to search tasks. Across all participants, the four tasks were performed for each image. While each individual participant solved all four tasks, only two of the tasks were solved for the same image in a randomized order.

The experimental data were used to explore how different task instructions influence model parameters. It is important to note that our approach required that time-

ordered scan paths for each trial are available, i.e., a sequence of N fixations, $\mathcal{F}_N = \{f_1, f_2, ..., f_N\}$, to evaluate the model. Each fixation i is a combination of fixation location x_i and fixation duration T_i. For the scan path $\mathcal{F}_N$, the log-likelihood is computed using Eq. (3.13). In order to limit the variability in scan path lengths we limited the maximum number of fixations per trial to 20, i.e., we removed the last fixations of a trial where necessary.

Experimental data were split into a training and a test sets. For the fixation densities that serve as input saliency maps to the SceneWalk model we used different data for general and task specific densities. The task specific saliency maps were estimated from all fixation sequences obtained from the corresponding task condition. The general saliency maps were computed based on experimental data from a separate study in which the same images were shown in a free-viewing paradigm (Backhaus & Engbert, 2022b).

3.4 Results

The key motivation for the current study was a model-based analysis of the influence of task on viewing behavior in natural scenes. Results from statistical model inference may be investigated at three different levels. First, we analyze the parameter values (obtained from the training data), which translate into process assumptions as they possess specific interpretations in our mechanistic model. For example, numerical values must fall within a range that is defined by its interpretation. Second, the model likelihood for the test data set indicates the quality of the fit and will be used to compare model variants. Third, we compare model-generated data to experimental data. In Bayesian analysis, this step is termed posterior predictive checks (Schad et al., 2021). Related analyses are highly indicative of which behavior the model captures well and which aspects might be caused by yet unidentified mechanisms. Capturing interindividual differences will be an important criterion for our model. The workflow for our dynamical modeling study is summarized in Figure B.1 in Appendix Section B.3.

3.4.1 Parameter estimation

In order to fit the parameters of the model to the task-dependent scene-viewing study (Backhaus et al., 2020), we implemented a Bayesian workflow as proposed by Schad et al. (2021), for which the likelihood computation for each scan path, Eq. (3.16), is an essential prerequisite. For MCMC sampling we used the PyDREAM implementation (Shockley et al., 2018) of the DREAM(ZS) algorithm (Laloy & Vrugt, 2012). Based on priors for model parameters (see Table 3.1) informed by previous studies (Schütt et al., 2017; Schwetlick, Rothkegel, Trukenbrod, et al., 2020b), PyDREAM generates samples converging to the posterior distribution over the model parameter space. We ran 3 chains of 20,000 iterations for 9 parameters for each of 32 participants in each of 4 tasks. These numerical computations were carried out for both the model variant

with task-dependent saliency input maps and for the model variant with one general-purpose saliency input for all tasks. Thus, we report data from 256 model fits of 9 parameters each. As suggested by Vrugt and Braak (2011) we verified the convergence of the estimation using the Gelman-Rubin $\hat{R}$ statistic[5] (see Appendix Section B.4, Figure B.2).

Due to a combination of the number of models, model parameters, and number of iterations for each scan path, we conducted parameter estimations on a medium-size multi-core system. The sequential nature of scan paths computations allows parallelization of the iterations between scan paths but not within. One likelihood evaluation, i.e., one iteration in the MCMC sampling algorithm, can be computed within about 10 seconds. One model (out of the total 256), using 28 CPUs, with three parallel chains of 20,000 iterations required an approximate computing time of 55 hours.

In the Bayesian approach, the posterior density contains all information about the model parameters. Figure 3.3 shows the marginal posteriors of all estimated model parameters (Tab. 3.1) for task-specific saliency maps, as this is our baseline model. The parameters in most cases converge to a distinct posterior distribution, which encode individual differences. As an example, σ_A and σ_F should be noted as parameters where the differences between the participants are explained as differences in attentional span variability.

We now discuss important effects of the task on the SceneWalk model parameters using task-specific and general saliency model variants, as reported in Table 3.2. Note that in Figure 3.3 as well as in Table 3.2 we report general population-level trends, although the models were fitted individually for each subject and task. To give overall interpretation of results, we averaged the marginal posteriors and report maximum posterior density measures of this average. It is important to note that these approximations are not equivalent to parameters fitted generally to the whole population and disregard correlations between parameters. This measure is used solely descriptively; all further analyses and statistics were conducted using the full marginal posteriors for each model fits.

The parameters σ_A and σ_F represent the sizes of the Gaussian-shaped inputs for the activation and inhibition streams, respectively. Both are smaller in the Count conditions than in the Guess conditions, indicating a more localized focus in the Count conditions. In fact the Count Animals condition is characterized by the smallest values for both parameters. Locations of animals in the photographs are very diverse, requiring detailed inspection and making the task the most difficult of the four.

Parameter ζ is the noise term. It is larger in the Count conditions, indicating that the data is less predictable in those conditions than in the Guess conditions (note that ζ is plotted as $-\log(\zeta)$ and is therefore negative; smaller values have larger negative values). This could again be interpreted as the result of the more directed viewing behavior in count tasks.

5 Full details of the implementation of the model, the inference, and a variety of checks and to ensure correct behavior are included in the OSF repository.

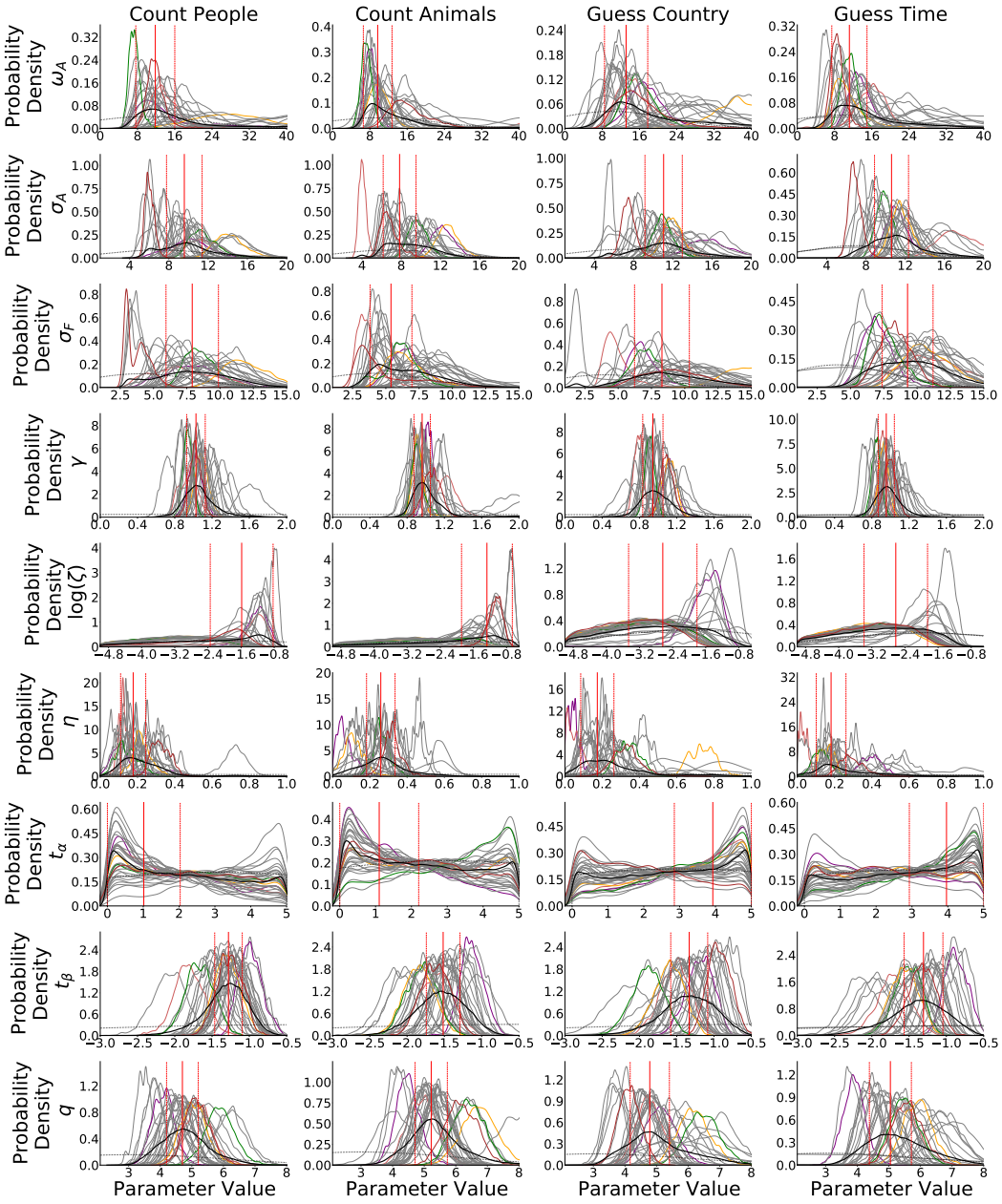

Figure 3.3 Marginal posteriors for all estimated model parameters across the four tasks. The panels visualize the marginal posteriors for all estimated model parameters using task-specific saliency maps. The columns indicate the four tasks; rows represent the 9 estimated model parameters. Each grey line is one subject. The colored lines correspond to data from arbitrarily selected participants, so that results for some participants can be compared across different parameters and tasks. The black lines represent averages over all participants, i.e., a kernel density estimate computed jointly for the samples of all models which shows the trend and spread of parameter values. The dotted lines visualize the prior distributions. Vertical red lines mark the 50% highest posterior density interval.

Model	Parameter	Count Animals		Count People		Guess Country		Guess Time	
		+/-	Point Estimate	+/-	Point Estimate	+/-	Point Estimate	+/-	Point Estimate
General Saliency	γ	0.099	0.913	0.081	1.053	0.143	0.932	0.091	0.938
General Saliency	ω_A	2.457	8.378	2.863	8.918	4.774	12.361	4.280	10.927
General Saliency	η	0.083	0.253	0.060	0.179	0.089	0.170	0.084	0.182
General Saliency	σ_A	1.710	6.913	1.744	9.005	1.936	10.506	1.632	10.224
General Saliency	σ_F	1.292	5.049	1.790	8.146	2.523	8.437	1.911	9.116
General Saliency	t_α	1.036	3.964	1.185	3.815	1.171	3.829	1.019	3.981
General Saliency	t_β	0.220	-1.350	0.174	-1.109	0.240	-1.352	0.259	-1.307
General Saliency	q	0.498	5.053	0.477	4.475	0.618	4.757	0.664	5.010
General Saliency	$log(\zeta)$	0.415	-1.122	0.492	-1.208	0.842	-1.861	0.738	-1.998
Task-specific Saliency	γ	0.088	0.961	0.100	1.027	0.109	0.943	0.087	0.955
Task-specific Saliency	ω_A	3.080	9.680	4.187	11.815	4.642	13.129	3.779	11.184
Task-specific Saliency	η	0.077	0.259	0.067	0.177	0.089	0.174	0.080	0.181
Task-specific Saliency	σ_A	1.674	7.832	1.814	9.573	1.903	11.060	1.733	10.608
Task-specific Saliency	σ_F	1.573	5.387	1.970	7.910	2.058	8.287	1.916	9.283
Task-specific Saliency	t_α	1.101	1.101	1.015	1.015	1.074	3.926	1.037	3.963
Task-specific Saliency	t_β	0.226	-1.517	0.185	-1.280	0.248	-1.333	0.261	-1.307
Task-specific Saliency	q	0.511	5.222	0.501	4.701	0.621	4.786	0.665	5.047
Task-specific Saliency	$log(\zeta)$	0.613	-1.273	0.759	-1.585	0.826	-2.637	0.767	-2.622

Table 3.2 Point estimates for each parameter by task. The reported point estimates for each parameter and model are the center of the 50% maximum posterior density interval, averaged over subjects.

The coupling of local saliency (or empirical fixation density) and mean fixation duration is an important new component of the SceneWalk model. For all of the four task conditions, the coupling parameter t_β, Eq. (3.12) turns out to be negative with zero outside the credibility interval. Thus, for higher saliency at position x_i compared to position x_j, i.e., $0 < s(x_j) < s(x_i) < 1$, we have $\log s(x_j) < \log s(x_i) < 0$. Since $t_\beta < 0$, the rate parameter b will be larger at position x_j compared to position x_i. Finally, since the mean fixation duration $\mu = q/b$, we obtain a longer mean fixation duration at the high-saliency position x_i compared to the low-saliency position x_j. Therefore, in our model, image patches of higher saliency will be fixated longer on average. This is in good agreement with the results obtained for the LATEST model, where the decision rate is negatively correlated with saliency (Tatler et al., 2017).

3.4.2 Likelihood for general versus task-specific saliency

The model likelihood informs about the overall adequacy of the model for explaining the experimental data. An important theoretical question is related to the relative performances of the model variants with general and task-specific saliency maps. For example, are task-specific effects primarily due to task-specific saliency maps or can we find task-specific parameters in scan path generating processes? To answer these questions, we fitted two model variants, one model variant where the input saliency map was computed from the experimental fixation density of a free viewing task and another model variant where each fixation sequence was obtained from task-specific experimental data.

Mathematically, the new model offers an interesting perspective with respect to saccade timing and spatial target selection. We introduced the extended model with an explicit saccade timing mechanism. Based on the model formulation, we showed that the likelihood function can be decomposed into a spatial and a temporal component, Eq. (3.15). Since saccade timing and spatial target selection should be looked upon as partially independent systems (Findlay & Walker, 1999a), we investigate these two likelihood components separately (see Fig. 3.4).

Figure 3.4 reports spatial and temporal likelihood components based on the test data. We represent the values as information gain in bit per fixation compared to a random null model. The spatial null model is random selection of points from the grid according to the assumption of complete spatial randomness (Illian et al., 2008) with log-likelihood $\log(\frac{1}{128^2})$. The corresponding temporal null model is based on the assumption of a constant probability for saccade onset, which gives a Poissonian waiting time distribution with a rate λ corresponding to the average number of fixations per trial found in the data.

We observe that the models with both task-specific saliency and general saliency can be fitted equally well with respect to the temporal likelihood (Figure 3.4a). Even though t_β the parameter that couples saliency and durations, is non-zero, the added information of task-specific saliency maps do not improve the temporal likelihood of the model.

In the spatial likelihood, the task-specific saliency maps generate an advantage for

the model (Figure 3.4b), which can be expected, since the task-specific saliency map is based on specific experimental fixation densities. Sampling from the saliency map is often used to measure for model performance in static models. It is interesting to note that both SceneWalk model versions significantly outperform random sampling from the saliency map. Furthermore, the extended SceneWalk model that uses only general saliency maps outperforms a model that samples randomly from the task-specific saliency map. This result suggests that the dynamical mechanisms in fixation selection are as task-specific as the saliency map.

To analyze the modeling results statistically, we calculated a set of linear mixed models (LMMs) including three fixed-effects (Bates et al., 2015): Factor Model (Density Sampling vs. SceneWalk), Factor Saliency (general vs. task specific saliency), and the interaction of both (i.e., the interaction Model:Saliency). As variance components, we estimated a separate intercept for each subject and for each image. For this analysis we consider values $|t| > 2$ as significant, which produces the following result. Firstly, we find an effect for Factor Model $\beta_{Model} = 0.34\ bit/fix$ (using SceneWalk improves the information gain by $0.34\ bit/fix$ over Density Sampling). Secondly, we find an effect for Factor Saliency $\beta_{Saliency} = 0.25\ bit/fix$ (task-specific saliencies improves the information gain by $0.25\ bit/fix$ over general saliencies). Lastly there is no effect for the interaction. Inspection of residuals of the model fit identified 11 outliers out of the total of 1700 data points. We tested a refitted model without the outliers, which did not affect the profile of significant effects. In an additional LMM, we calculated a treatment contrast to statistically validate the difference between 'Task-specific Saliency–Density Sampling' vs. 'General Saliency–SceneWalk'. All comparisons with our baseline 'Task-specific Saliency–Density Sampling' turned out to be significant. The comparison of our main interest revealed a significant difference of $\beta = 0.09\ bit/fix$ with $t = 2.80$. This significance is compatible with the idea that task dependent scan path dynamics contribute reliably to the model beyond the static task differences (i.e., fixation densities).

3.4.3 Posterior predictive checks: Fitting scan path statistics

Posterior predictive checks refers to the investigation of data generated by the model after parameters have been identified. Model-simulated data may be compared to experimental data in a variety of metrics beyond the model likelihood. As has been shown in previous work (Schwetlick, Rothkegel, Trukenbrod, et al., 2020b), the SceneWalk model is capable of fitting a variety of metrics of scan path dynamics, beyond mean fixation durations and mean saccade amplitudes as well as their distributions. One important measure of scan path generation is the distribution of turning angles, specifically as a function of saccade amplitude and fixation duration. Here, the posterior predictive checks are important in order to ascertain that our changes to the model architecture did not degrade the fit of scan path statistics with respect to the previous model version (Schwetlick, Rothkegel, Trukenbrod, et al., 2020b). In Figure 3.6 we show that the model fits achieved for this data are well-fitted to the experimental distribution from the test data.

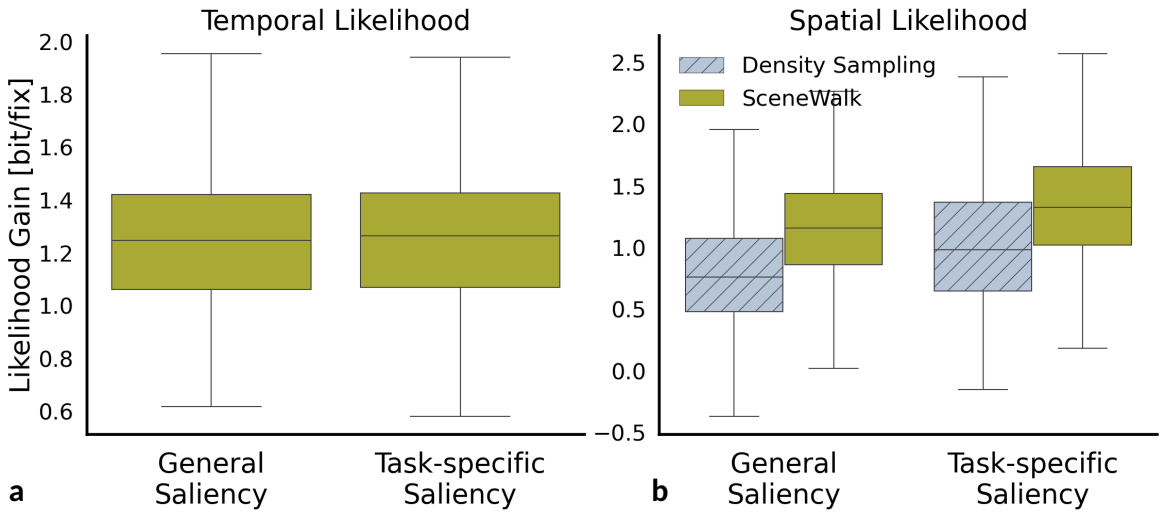

Figure 3.4 Spatial and temporal likelihoods for model variants. Comparison of the model likelihood gain for general and task-specific model variants. **(a)** The temporal likelihood gain of the SceneWalk model is computed as the difference between the model likelihood and a statistical model (Poisson waiting time distribution). **(b)** The spatial likelihood gain of the SceneWalk model is obtained as the difference to complete spatial randomness. As a baseline model, we compare the numerical results against a density sampling model (grey, hatched bars) without any dynamics. The combination of the SceneWalk model and the general saliency model outperforms the task-specific density sampling. To improve the visibility of the effects, we omit outlier points in the box plot.The full results of the linear mixed model are supplied in the Appendix (Table B.1).

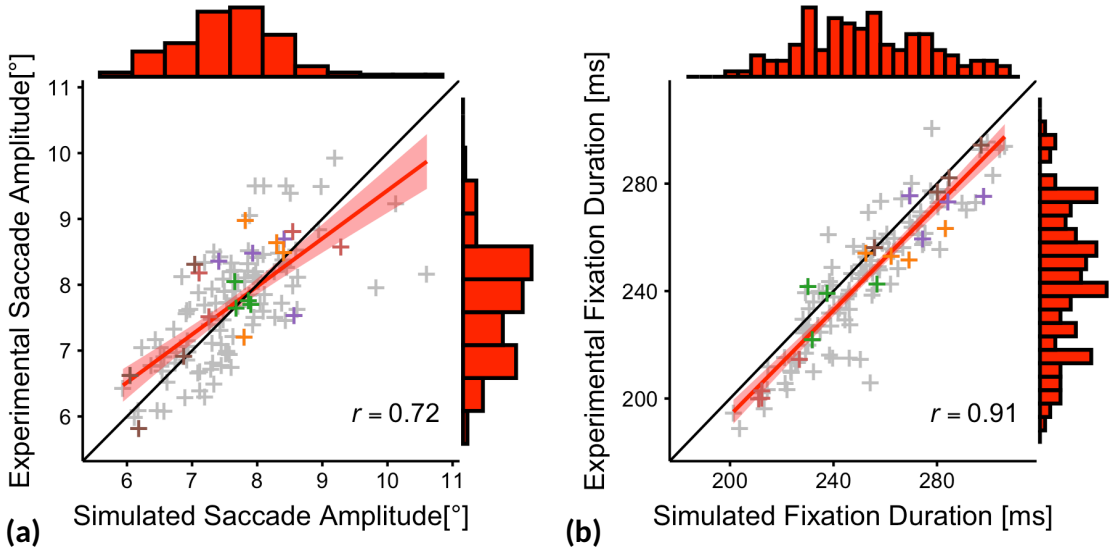

Figure 3.5 Correlation between experimental and model-generated saccade statistics. **(a)** Mean saccade amplitude with experimental data plotted along the vertical axis and simulated data plotted along the horizontal axis. Each grey cross ("+") represents experimental and simulated data for an individual observer in a specific task condition. The colored crosses indicate individual observers with the same color mapping as in Figure 3.3. The red line gives the regression line. Histograms at the top and right side of the panel visualize the distributions of saccade amplitudes for simulated and experimental data, respectively. **(b)** The analogous plot for fixation durations of individual observers in specific task conditions.

Generally, posterior predictive checks are necessary for investigating the presence of important characteristics of the experiments in the model-generated data. The above examples for the influences of saccadic turning angle on saccade amplitude and fixation duration are crucial for any scan path model. It is important to note that fitting the model based on the likelihood function and without a consideration of specific ad-hoc metrics produces the correct behavior reliably. As an additional constraint, parameter estimation presented in the current study had access to greatly reduced amount of data compared to previous studies, due to the fact that the model was fit to individual observers on the training subset of a data set with limited trials. We interpret the stable emergence of the critical characteristics of behavior in spite of this as an assurance that the fitting procedure was successful and the experimental data support our model hypotheses strongly.

Figure 3.5 shows the correlation of experimental and simulated saccade amplitudes as well as fixation durations for each subject and task, i.e., for each individual set of model parameters. We find a high correlation for both measures, indicating that the model reproduces important summary statistics in the data. Moreover this plot illustrates the way in which the model is able to capture interindividual differences. A model fit to a particular participant who experimentally tends to produce longer than average saccades, will also produce longer saccades when simulating data and vice

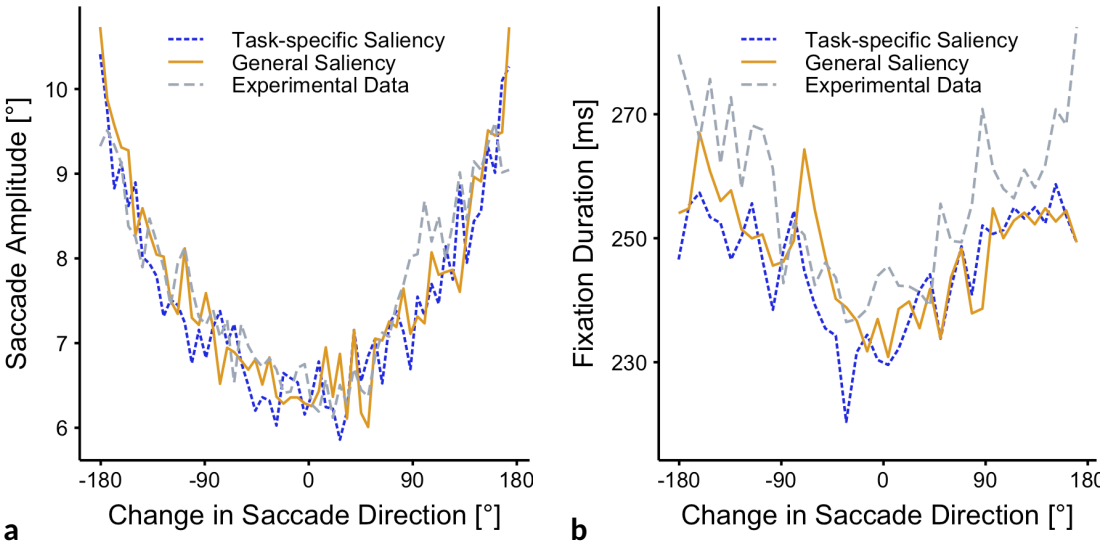

Figure 3.6 Saccade turning angle as function of saccade amplitude and fixation duration. (a) Plot of the saccade amplitude as a function of the change in saccade direction (i.e., saccade turning angle). Averages for model variants with general and task-specific saliency (colors) are compared to experimental data (dashed line). **(b)** Same plot for fixation duration as the dependent measure.

versa. The same is true for fixation durations. These correlations are an important measure for the sensitivity of the model with respect to interindividual variation. Differences between fits (as shown in Figure 3.3) are not caused by noise or fitting errors but are explaining between-subject variance.

3.4.4 Statistical analysis of model parameters from posteriors

Since in the Bayesian approach we obtain the posterior density over the space of model parameters as a result of model inference, we will be able to run a detailed statistical analysis of the parameter variations across tasks and individual observers. We used linear mixed models (LMMs) to analyze the differences between tasks for each parameter (Bates et al., 2015). As before, we analyzed both models, i.e., the general saliency and the task-specific saliency models.

For the statistical analyses, we sampled parameters from the full posterior density. We ensured the samples were independent by thinning the posterior to every 100th sample and checked statistical independence by analysis of the autocorrelation function. A separate LMM was calculated for each parameter and both the general and specific task models. The results of these analyses are shown in Figure 3.7.

The fixed effect structure is taken from Backhaus et al. (2020), where contrast coding follows the approach of Schad et al. (2020). We chose a random effect structure with a varying intercept and a varying slope for each contrast by every subject. We did not include image as a factor in the random effect part of the LMM, as we did not model parameters separately for every image. The resulting model, presented in the model notation of the lme4 R package (Bates et al., 2015; R Core Team, 2019), can

be written as

$$\mathrm{DV} \sim 1 + FGC + FC + FG + (1 + FGC + FC + FG||\mathrm{subject}) \,, \tag{3.17}$$

where DV represents the dependent variable. The symbol "1" represents the model's intercept, FGC denotes the first contrast of both Guess against both Count tasks; FC is the second contrast of Count Animals against Count People conditions; FG denotes the third contrast of Guess Time against Guess Country tasks. The correlations of random effects are not included in the model, which is represented by the double bar sign || in the formula.

An important requirement of LMMs is that the residuals are normally distributed. We checked the distributions and calculated an optimal λ-coefficient via the Box-Cox power transform (Box & Cox, 1964) to re-adjust the experimental data. Even after transform, model residuals of some parameter estimations deviated from a normal distribution. However, Schielzeth et al. (2020) addressed the consequences of violations in distributional assumptions and identified only slightly upwards biases in estimates of varying effect variance. Thus, we expect our results to be reliable in general.

First, referencing Figure 3.7, we compare fixed effect parameter estimations within the task-specific saliency variation (blue bars). With this model we make the assumption that saliency of image features changes in response to task and ask the question of whether this change in weighting is sufficient to explain the change in behavior. In our analysis, we find differences in parameter values between the two task groups (Guess and Count) for the parameters σ_A, σ_F, γ, ζ and t_α, t_β, q. Above, we qualitatively described the parameters σ_A, σ_F, referring to the attentional and inhibitory span, as well as the noise parameter ζ and timing parameters t_α, t_β, and q. In the task-specific saliency model we also find significant differences between Guess and Count tasks for γ.

The parameter γ controls the weighting of the selection map (priority map), making it more or less deterministic. Large values of γ lead to steeper peaks in the priority map and thus the target selection is more deterministic. Here, we find that count tasks lead to larger values of γ than guess tasks. We relate this finding to the task demands. The object search behavior needed for the Count task, particularly when given a task-specific saliency, is strongly focused on specific targets. The model therefore emphasizes peaks in the selection map, driving more precise and focused target selection by a higher value of the exponent γ compared to guess tasks.

Second, we compare within the task groups. As reported by Backhaus et al. (2020), the two Count conditions themselves evoke different behavior. Searching for animals is a more general tasks (they could be any species, so conceivably found on land or in the air or camouflaged) whereas counting humans is more predictable. Therefore, the difference between these two tasks also caused significant differences in the parameter estimates, specifically for parameters σ_F, t_β, and q. The model parameter σ_F, which is responsible for the size of the inhibitory fixation tagging mechanism, is smaller in the Count Animals condition. We interpret this finding by assuming that more local inhibition is particularly important for counting animals to permit finely guided

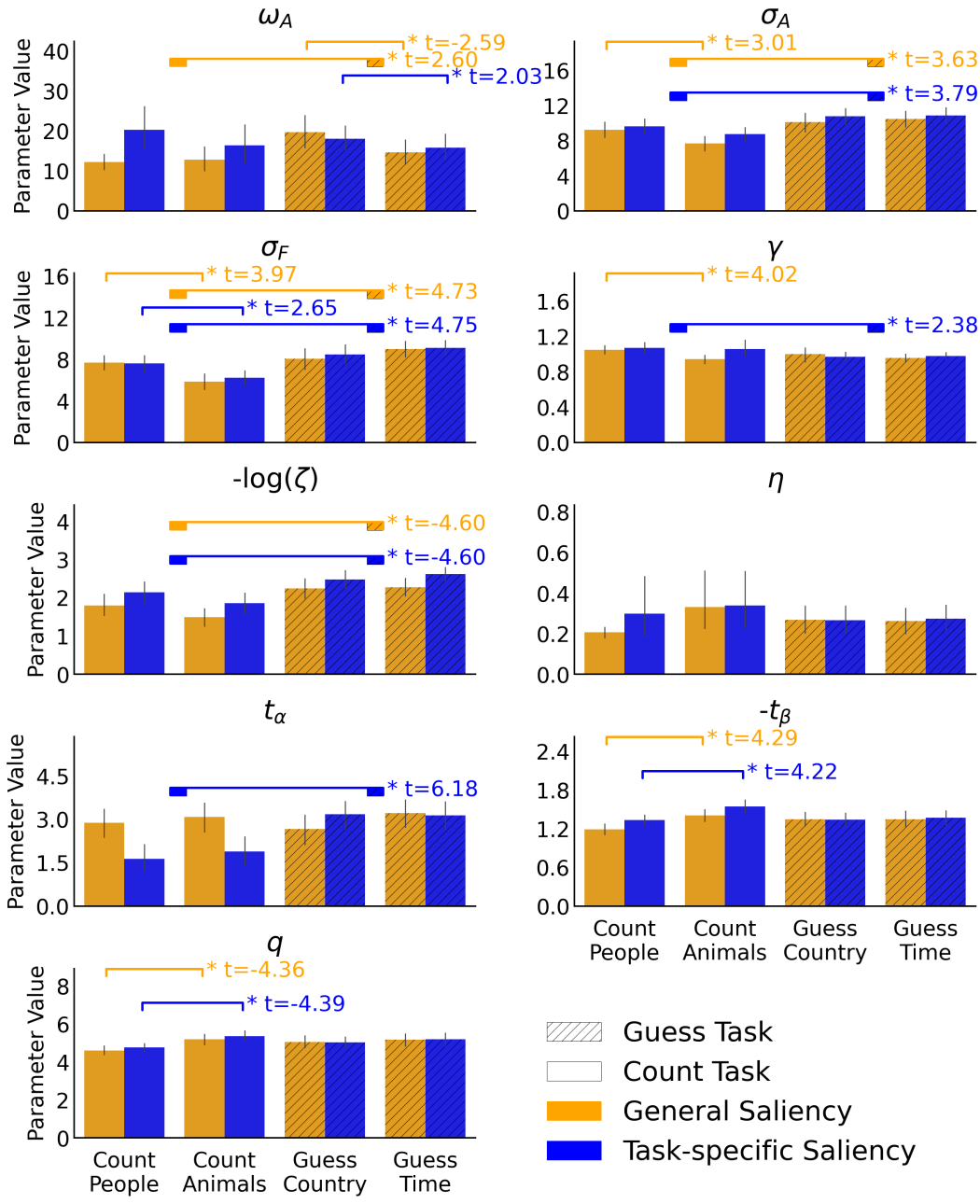

Figure 3.7 Comparison of parameter estimates between models and the different task conditions. The orange bars refer to the general saliency model; blue to the task specific model. The hatching highlights the two Guess tasks. Horizontal lines above the bars show the significant fixed effects as found by a mixed linear model.

refixations that might be necessary for counting densely packed scene content.

The saccade timing parameters t_β and q are also significantly different between the two counting tasks. Specifically, parameter t_β determines the influence of the saliency on the duration (the more negative t_β the stronger the influence of saliency on duration; see discussion in the section on parameter inference). Beyond the fact that the model parameters t_β and p reproduce experimental effects on the difference between the two counting tasks, we would also like to point out that saliency maps are more driven by people than by animal locations. When the task involves searching for people and people cause high saliency, it is most likely to find the search target in high saliency regions. The value of saliency thus influences fixation duration more strongly than in the Count Animals task—a trend that is visible in both fitted parameters t_β and p.

In the next step we investigated the parameter differences when the model was given the same, general saliency map for each task. In this condition too, we find differences between the task groups. Because the saliency itself has smaller explanatory value, the parameters of the SceneWalk model take on a more cogent role. In addition to the significances of the task-specific saliency model described above, we also find significant differences between the task groups in the parameter ω_A. This parameter specifies the speed of the activation decay. Here we find significantly slower decay for Count tasks than for Guess tasks in the absence of task-specific information. We suggest that this may be the case because it is more directly useful in search tasks to keep track of previous locations and significant areas.

The contrast defining the difference between the Count Animals and the Count People conditions is, as in the original analysis by Backhaus et al. (2020), also significant in some parameters: σ_F, γ, and η. Parameter σ_F, the size of the inhibition Gaussian is smaller for Count Animals condition. This may reflect the size of the objects that are typically being counted. The greater size of γ and smaller size of η (the length of the post-saccadic shift) in the Count People condition may be related to similar factors of the size and typical locations of the searched objects.

3.4.5 Statistical analysis of scan path statistics

In the previous section, we investigated model-based parameter variations across tasks and observers, which focussed on the models and the meanings of its parameters. Here, we switch to analyzes of the data, where we compare experimental and model-generated data with respect to scan path statistics.

Based on the estimated parameters per participant, we generated scan paths using the SceneWalk model. These simulated data will be compared to experimental scan paths to investigate whether the statistics of behavior and task differences are reproduced by the model. In the experimental study by Backhaus et al. (2020), the authors investigated how various scan path statistics, such as fixation duration or saccade amplitude vary with task. Tasks that can be roughly characterized as less-constrained free viewing tasks (here: Guess tasks) produce longer saccade amplitudes and longer fixation durations than tasks with a clear search component (here: Count tasks).

Previous research has also shown that saccade amplitude and attentional span are

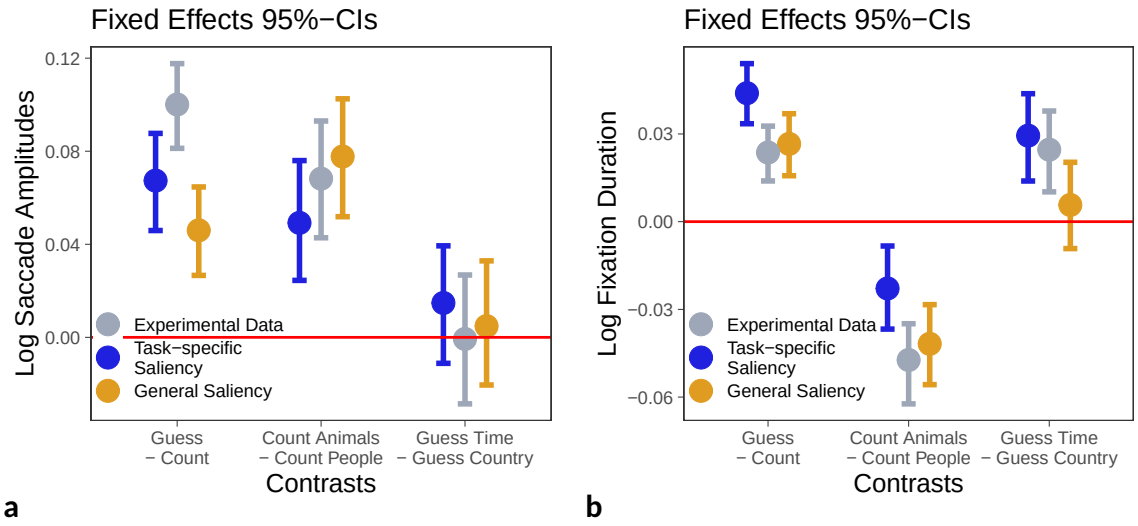

Figure 3.8 Comparison of fixed-effects from linear mixed model analysis. (a) Estimated fixed effects comparison for saccade amplitude analyses. **(b)** Estimated fixed effects comparison for fixation duration analyses. In both panels, grey color shows the LMM estimates for experimental data, orange color shows the general saliency model, and blue color represents the task-specific saliency model. The horizontal red line marks the zero value, at which there are no differences in the specified contrasts. Confidence intervals around the estimated effects are the bootstrapped shortest 95% intervals.

related and saccade amplitudes tend to be smaller in search tasks (Trukenbrod et al., 2019). We conducted the same linear mixed model analyses with the original contrast coding and fixed effect structure reported by Backhaus et al. (2020), but a reduced random effect structure with only varying intercepts for subjects and images (i.e., no varying slopes both on the simulated data and on the experimental data). The resulting model in the model notation of the lme4 R package (Bates et al., 2015; R Core Team, 2019), as well as an overview of LMM structures may be found in the Appendix Table B5.

We found that almost the same contrasts turned out to be significant and the estimated values are in good agreement in all cases (Figure 3.8). Results for saccade amplitudes are reported in Table B3; for fixation durations please consult Table B4. The only exception is between general saliency, task-specific saliency, and experimental data for the contrast that captures the difference between the two Guess tasks for fixation duration. Specifically, this contrast is significantly positive for experimental and task-specific saliency data. The estimate for the simulated data with a general saliency map, however, can vary to values below zero. Note that the model's responses are slightly muted as compared to the human scan paths.

3.5 Discussion

Visual exploration of natural scenes depends on the given objective. This has been noted since the beginnings of vision science (Yarbus, 1967). Based on the advances in modeling of visual attention (Itti & Koch, 2001) and eye-movement control (Schwetlick, Rothkegel, Trukenbrod, et al., 2020b; Tatler et al., 2017), we investigated the performance of a computational model of scan path generation for an experiment in scene viewing over four different tasks. We extended the SceneWalk model (Engbert, Trukenbrod, et al., 2015; Schwetlick, Rothkegel, Trukenbrod, et al., 2020b) by an explicit saccade timing mechanism and implemented a fully Bayesian framework for dynamical, process-oriented modeling (Schütt et al., 2017). Specifically, in this approach, it is possible to estimate model parameters for individual human observers. Thus, in posterior predictive checks, we were able to carry out a statistical analysis of individual differences across tasks. As a result, we found evidence for specific adaptations of model parameters to task constraints. The extended SceneWalk model reproduces task-effects, individual differences across tasks, and demonstrates an overall advantage for model variants with task-specific saliency maps.

Overall, our findings suggest that parameters in the generation of scan paths are as highly adaptive to task requirements as are saliency maps. First, given a specific task, human observers seem to adjust the control of saccade dynamics. This is psychologically plausible, since, for example, stronger inhibition of return and smaller saccade amplitudes might contribute to an effective strategy for fine-grained search behavior compared to a less-constrained free viewing task. Second, it is also psychologically plausible that the saliency of certain object features in scene changes with the requirements of the task. Looking for a specific object may result in a strategy of ignoring all features that are unlikely to be associated with that object.

3.5.1 Dynamical modeling of eye-movement control

Our current results are an example for process-oriented, dynamical modeling as tool not only for predicting human behavior (Engbert et al., 2022), but also for identifying gaps in our understanding. Over the last decade, major advances were related to model of visual attention in scene viewing, with the time-independent 2D fixation density as the modeling target (Itti & Koch, 2001; Koch & Ullman, 1985; Kümmerer et al., 2015; Pan et al., 2016). Recently, the interest is growing in predicting time-dependent series of fixations, both in the field of vision science (Engbert et al., 2005; Le Meur & Liu, 2015; Tatler et al., 2017) and in the context of deep learning (Kerkouri et al., 2021; Kümmerer & Bethge, 2021). In our process-oriented approach the SceneWalk model implements specific mechanisms inspired by successful experimental research such as inhibition of return (Klein, 2000; Klein & MacInnes, 1999). Interestingly, in our model inhibitory tagging is modulated by task, with a smaller spatial size of the inhibitory tagging parameter for counting compared to guess tasks—a finding that underlines the flexibility of the contributing attentional processes in eye-movement control.

An important advantage of the process-based approach over more data-inspired

models (Le Meur & Liu, 2015) or deep learning neural networks (Kümmerer, Theis, et al., 2014) is that there are fewer model parameters in process-based models, which have a clear interpretation with respect to their function in the control of eye movements. Thus, process-oriented models provide insights into how well our current understanding describes the process. Posterior predictive checks, i.e., the comparison of simulated and experimental data along a variety of metrics reveals the gaps between what is implemented in the model and the underlying process. The addition of the new timing mechanism in this work is an example of applying this approach. It is inspired by assumptions from the literature (Tatler et al., 2017), is confirmed by the estimated parameters (the best fit value for the coupling parameter t_β is non-zero, indicating that spatial and temporal components are linked), and is validated by posterior predictive checks. Finally, we applied our model and our framework for parameter inference to the estimation inter-subject variability and inter-task variance in scan paths.

3.5.2 Model adaptivity: task-specific model parameters

The SceneWalk model produces specific, systematically different parameter estimations when fit on data from a range of tasks. Using considerable computational resources we conducted separate model fitting procedures for each subject and each task for two model versions. The parameter estimation successfully found an informative posterior distribution in the majority of cases. These marginal parameter posteriors reveal pronounced differences in value for the different tasks and subjects. The success of the estimation is worth noting particularly because the amount of data available for the number of estimations was comparatively small. This work contributes insights into the relevance of task and interindividual differences for the process of attention selection. In the next paragraphs we will discuss the parameter differences in detail.

The two most straight-forwardly interpretable parameters in the model are attention span σ_A and the inhibition size σ_F. The estimated parameter values for both are larger in Guess tasks than in Count tasks. We propose to interpret this in the following way. A reduced attentional span enables a detailed inspection of small areas. This is consistent with the finding that search tasks elicit more and shorter saccades. The length of the saccades and the estimated attentional span in our model are highly correlated. For free viewing tasks, a broader attentional span is useful as it allows the viewer to take a wider perspective and take into account more features, but with less detail. In fact we find that the smallest attentional span is found in the count animals condition. This is also the most detail-oriented and difficult task. The inhibition size is also smaller for count conditions. We propose that this is partially a direct result of the amount of inhibition needed to counteract the activation and partially due to a more precise tagging of already-viewed locations. Thus, the parameters reflect the influence of task on spatial gaze statistics (Backhaus et al., 2020; Mills et al., 2011).

The parameter ω_A regulates the speed of the activation decay. The speed of the two streams, activation and inhibition, is separated by an order of magnitude ($\omega_A/\omega_F = 10$). We find a slight difference between Count and Guess tasks for the parameter ω_A when we fit with one general saliency. Specifically, Count tasks have a systematically

slower decay than Guess tasks. We would like to put forward the interpretation that in the case of search tasks the past positions retain more importance. The searcher needs to keep track of already fixated or found objects as well as inhibit discovered distractors. However, we find this effect only when the input saliency is general; no such difference emerges when the saliency map is more informative and task-specific. A possible explanation is that the information which is available longer due to slower decay is related to the information which, in the other case, is present in task-specific saliency maps. That is, the information can either be encoded in the saliency input map or can be accounted for by slower decay(ω_A). In one case the model has to build that representation itself (general case) and in the other it does not need to as the information is in the input saliency (specific case).

3.5.3 Temporal control of saccades

In this study, we we provided an important extension of the SceneWalk model to temporal control of saccades. It is important to note the earlier version of the model (Schwetlick, Rothkegel, Trukenbrod, et al., 2020b) included saccade-related modulations of fixation duration, but not an explicit timing mechanism. The explicit saccade timing enables the model to make predictions not just for the spatial selection of fixation locations and the interaction with fixations, but also for modeling task-dependent, strategic effects in mean fixation durations.

The new timing mechanism introduces additional variability in the coupling between spatial and temporal selection. The control of fixation durations in scene-viewing were studied earlier based on explicit timing mechanisms (Laubrock et al., 2013; Nuthmann & Henderson, 2010). Most recently, the LATEST model combined temporal with spatial aspects of saccade generation (Tatler et al., 2017). While the dynamical part of the LATEST model is limited to the saccade timing, it motivated the integration of a timing component to our fully dynamical framework (Engbert, Trukenbrod, et al., 2015; Schwetlick, Rothkegel, Trukenbrod, et al., 2020b). We successfully implemented a coupling of the local saliency at the current fixation location to mean and variance of the saccade timer. The prior for the spatiotemporal coupling parameter t_β included the option for this magnitude to be zero, i.e., to infer that saliency has no influence on duration, effectively decoupling the two components. In accordance with our expectations and with the results of the LATEST model (Tatler et al., 2017), the credibility interval t_β did not include zero (numerically, the mean is between 1 and 1.5). Thus, we obtained clear evidence for longer average fixation durations at image patches with higher saliency compared to region of lower saliency.

The likelihood function plays an important role for combined modeling of fixation durations and fixation locations (e.g., Engbert et al., 2022; Schütt et al., 2017). To our knowledge, the first study using spatiotemporal likelihood inference in scene viewing was published by Kucharsky et al. (2021), in line with conceptual work for eye movements in reading by Seelig et al. (2020). Kucharsky et al. (2021)'s WALD-EM model combines a standard information accumulation process for saccade timing with a spatial component. Similar to our results, WALD-EM was demonstrated to successfully

reproduce several key aspects of eye-movements statistics including interindividual differences. Different from WALD-EM, our model includes biologically motivated, perisaccadic attentional processes around the time of saccade to reproduce several experimentally observed qualitative phenomena such as couplings of fixation durations and turning angles (Schwetlick, Rothkegel, Trukenbrod, et al., 2020b). Thus, our approach implements a more complicated internal model structure. Both Kucharsky et al. (2021)'s WALD-EM and our SceneWalk model demonstrate the superiority of parameter inference based on a spatiotemporal likelihood function.

3.5.4 Interindividual differences in viewing behavior

An important step forward in dynamical modeling of individual viewing behavior was achieved by the likelihood-based framework for parameter inference. Experimentally, it is well known that saccade statistics and visual attention show marked interindividual differences (Kliegl, 2010; Makowski et al., 2020). In the past, modeling of an individual observer's behavior was out of reach, since model fitting based on ad-hoc statistics required an amount of data that was typically not provided by experimental studies. As a consequence, model parameters were estimated for data pooled over all of the participants of an experimental study, which precluded modeling of interindividual differences.

As parameter fitting algorithms have improved, it has become possible to reduce the amount of data needed. With the likelihood function available for the SceneWalk model, parameters could be inferred from experimental data on a single-subject level (Schwetlick, Rothkegel, Trukenbrod, et al., 2020b). Using the task specific data sets in this study, we had to further reduce the amount of data available to our fitting procedure. Fortunately, our MCMC implementation based on the DREAM(ZS) algorithm (Laloy & Vrugt, 2012) produced stable posteriors for each individual observer and across tasks.

3.5.5 General vs. task-specific saliency maps

As in our previous studies (Engbert, Trukenbrod, et al., 2015; Schütt et al., 2017; Schwetlick, Rothkegel, Trukenbrod, et al., 2020b), we focus on the investigation of dynamical principles of scan path generation. Therefore, we used experimental density maps as an upper bound for visual saliency models. Because of the available amount of observations, fixation-density maps could be produced from experimental data with specific task instructions.

One view of task differences in eye-movement control is that the differences mainly occur due to a saliency weighting of different aspects of the image. We might expect, therefore, when using task-specific saliency maps, our model of saccade dynamics to produce very similar parameter estimates for all tasks, since most of the variance in experimental data will be accounted for by the saliency maps. Interestingly, our analyses indicated that saccade dynamics strongly contribute to the adaptive behavior in response to task requirements. Model parameters, e.g., attentional and inhibitory

span (i.e., the sizes of the activation and inhibition Gaussians) or parameters related to temporal control of fixation durations turned out to be clearly different between the investigated tasks.

In addition to this task-specific saliency account, we also identified model parameters in a model variant where the visual saliency was derived from free viewing for all tasks, which we called the general saliency approach. Here, the underlying assumption is saliency is predominantly image dependent and does not change with task. The strongest version of this assumption implies that the observed variation in eye movement behavior is caused by the adjustment of the eye dynamics to task constraints. We found that the model of saccade dynamics still produces reasonable parameter estimates, however, the overall performance of the model was clearly weaker than for the model with task-specific saliency. One might argue that the psychologically plausible assumption would be that adaptation occurs in the saliency map as well as in the eye-movement dynamics. Nevertheless, it still seems very interesting that the model with general saliency outperforms density sampling from task-specific saliency maps. Thus, dynamics contribute significantly to task adaptation. A practical implication of this finding is that in a situation where only general saliency maps are available, adaptation of the eye-movement control system can significantly improve model predictions.

Modern models of visual saliency are usually evaluated with respect to scan paths generated by human observers, and, therefore, will contain both early saliency effect (Itti & Koch, 2001) and high-level influences from scene semantics (Henderson & Hayes, 2018). The same is true of the experimental fixation densities which are used in the SceneWalk simulations. Thus, the question of whether visual saliency is task dependent, is contingent upon the operational definition of saliency.

3.5.6 Model performance: posterior predictive checks

One of the key improvements presented in the study is the likelihood-based parameter inference for modeling individual viewing behavior (Engbert et al., 2022). While likelihood is mathematically rigorous, a maximum-likelihood model's performance can still be poor with respect to qualitative effects. Therefore, we carried out extensive posterior predictive checks, which demonstrated that our model reproduced many of the scan path statistics on the level of individual observers. Moreover, the model explained systematic differences of scan path statistics between the tasks found the in the underlying experimental study (Backhaus et al., 2020). As a dynamical and generative model, the SceneWalk model is capable of simulating scan paths given the estimated parameters and the saliency of an image. We simulated data for each observer and task as well as for both model versions based on general or task-dependent saliencies. Simulated data were compared to the experimental test data. The good agreement between the scan path statistics of simulated and experimental data is an essential component of a psychologically and biologically plausible model. Similarities and dissimilarities allow conclusions about which components of the process are well-captured by the model architecture and which still require explanation. In the current

study, we report good agreement between model-generated and experimental data. First, we confirm that the same general scan path statistics can be captured well with the extended model that includes temporal control of fixation duration compared to the latest version before this extension (Schwetlick, Rothkegel, Trukenbrod, et al., 2020b). Second, we compare whether the model captures the task differences found in the experimental data set (Backhaus et al., 2020). The experimental data indicated pronounced differences between the tasks in fixation duration and saccade statistics (e.g., amplitudes). Therefore, our results lend theory-based support to the idea that different viewing strategies are driven by saliency weighting, but also by dynamics of eye guidance.

Further analyses are required to test the model against additional experimental data sets covering a broader variation of task type. In the current work, the model's performance could be improved in view of the comparison to neural network models such as DeepGaze (Kümmerer, Theis, et al., 2014). While our model can be fitted to data from individual observers, interindividual variability will be overestimated due to differences in the convergence and identifiability of model parameters. Regularization by hierarchical modeling might be a solution here. Therefore, introducing hierarchical Bayesian dynamical modeling might be another big step forward for modeling individual observer's viewing behavior.

3.5.7 Evaluation of our preregistered hypotheses

Prior to conducting the current study, we preregistered our research plan including the main hypotheses (Schwetlick, Backhaus, & Engbert, 2020). The first two hypotheses in the preregistration concerned the task influence on attentional span and inhibitory fixation tagging. Specifically, we assumed that the attentional span would be larger in the Guess conditions, which we characterized to be similar to free viewing tasks. Previous research shows that saccade amplitude and attentional span are related and that saccade amplitudes tend to be smaller in Count tasks (Trukenbrod et al., 2019). The results from the estimation of parameter σ_A finds support for this hypothesis. Second, we proposed that inhibitory fixation tagging would be more important in search tasks. In fact we find that in count conditions the inhibitory tagging is more directed and less global, i.e. that the parameter σ_F is smaller, resulting in more specific inhibitory tagging of regions.

The third hypothesis concerns the decay of past states in the model. We expected for the Count conditions that the decay would be slower compared to the other tasks, since it might be more important to keep track of visited items. In accordance with this idea, we find slightly faster decay in Guess tasks than in Count tasks, as specified by ω_A in the general saliency model. In the appendix, we provide some more detailed summary and evaluation of our predictions and findings.

3.6 Conclusions

In this work we proposed an advanced model of eye-movement control with application to task-dependent viewing behavior. First, we extended a previous model to include temporal control of fixation durations and the interaction with spatial selection. Second, we applied rigorous statistical parameter inference that showed markedly different results across four different viewing tasks. These findings were corroborated by posterior predictive checks which indicated that these differences also manifest in data simulated by the model fits. Specifically, the model-simulated data reproduced the key scan path statistics found in experimental data. Thus, parameter inference yielded individual parameter estimates not only for tasks but also for each participant in the experimental data. We conclude that the SceneWalk model explained individual differences and task influences on behavior in a theoretically coherent framework.

3.7 Acknowledgements

This work was supported by a grant from Deutsche Forschungsgemeinschaft (DFG), Germany (SFB 1294, project no. 318763901). D.B. and R.E. received additional support via grant (EN 471/16-1, DFG).

4 Modeling Fixational Movement

Bayesian Dynamical Modeling of Fixational Eye Movements

Lisa Schwetlick, Sebastian Reich, Ralf Engbert
University of Potsdam

Abstract

Humans constantly move their eyes, even during visual fixations, where miniature (or fixational) eye movements are produced involuntarily. Fixational eye movements are composed of slow components (physiological drift and tremor) and fast microsaccades. The complex dynamics of physiological drift can be modeled qualitatively as a statistically self-avoiding random walk (SAW model, see Engbert et al., 2011). In this study, we implement a data assimilation approach for the SAW model to explain quantitative differences in experimental data obtained from high-resolution, video-based eye tracking. We present a likelihood function for the SAW model which allows us apply Bayesian parameter estimation at the level of individual human participants. Based on the model fits we find a relationship between the activation predicted by the SAW model and the occurrence of microsaccades. The latent model activation relative to microsaccade onsets and offsets using experimental data reveals evidence for a triggering mechanism for microsaccades. These findings suggest that the SAW model is capable of capturing individual differences and can serve as a tool for exploring the relationship between physiological drift and microsaccades as the two most important components of fixational eye movements. Our results contribute to the understanding of individual variability in microsaccade behaviors and the role of fixational eye movements in visual information processing.

Published on **ArXiv, 2023**

4.1 Introduction

The human eye is never truly at rest. At a macro-level the eyes move in a sequence of fixations and saccades (Schwetlick, Rothkegel, Trukenbrod, et al., 2020b), moving different aspects of the visual world into the receptor-dense center of the visual field. Despite the misleading term, even during fixations, the eyes are far from stationary (Ditchburn et al., 1959; Kowler, 2011; Martinez-Conde et al., 2004; Rucci & Victor, 2015). Microscopic fixational movements constantly shift the visual input over the receptors in the retina. Three components of fixational eye movement behavior are distinguished: a slow, meandering physiological drift, a high frequency tremor, and high velocity microsaccades (Alexander & Martinez-Conde, 2019). The comparatively small amplitude of tremor movement can not be resolved in current video-based eye tracking and is therefore neglected in the following.

The function of fixational eye movements, their underlying mechanisms, as well as the consequences for processing in the visual system have yet to be fully understood. While fixational eye movements are found ubiquitously across species and in all primates Ko et al., 2016, the generation of fixational eye movements varies between individuals and is highly characteristic (Cherici et al., 2012; Engbert & Kliegl, 2003, 2004; Poynter et al., 2013). In order to model the individually characteristic spatial statistics of physiological drift and the relationship to microsaccades we begin with a short discussion of the two movement components.

During physiological drift, the eyes move smoothly in a pattern that resembles Brownian motion over small time lags (Burak et al., 2010; Engbert, 2006; Pitkow et al., 2007), i.e., it meanders quasi-randomly, increasing variance over time. However, a more detailed analysis suggests that fixational eye movements represent an interesting example of fractional Brownian motion (Metzler & Klafter, 2000). A corresponding analysis can be carried out by computing the mean square displacement (MSD) at different time lags (Engbert & Kliegl, 2004; Herrmann et al., 2017). In Brownian motion, the MSD increases linearly with time lag. In fixational eye movements, however, a superdiffusive tendency is found over short time scales ($\lesssim$ 50 ms), which is also referred to as persistence. Over longer time scales ($\gtrsim$ 100 ms) physiological drift is found to be anti-persistent . This behavior can be interpreted by assuming that fixational eye movements maximize movement over short time scales to counteract retinal fatigue while reducing variance over longer time scale to maintain visual fixation on a intended region of interest or object (Engbert & Kliegl, 2004).

Microsaccades share their kinematic properties with their larger counterparts, such as acceleration profile, main sequence (i.e., the linear relationship between log amplitude and log peak velocity; see Bahill et al., 1975), and are generated by the same neural circuits in the brainstem (Rucci & Poletti, 2015). Microsaccades are distinguished mainly by their smaller amplitude, usually thresholded at $< 1^\circ$ (see Poletti and Rucci, 2016 for a review). Microsaccades occur at a rate which is highly variable between subjects (Engbert, 2006; Engbert & Kliegl, 2003). Attentional mechanisms also modulate microsaccades in both their rate, e.g., first, a reduced microsaccade rate following target onset, and then an increased rate (Engbert & Kliegl, 2003), and their

direction, e.g., the location of covert attention attracts microsaccades in endogeneous spatial cueing (Engbert & Kliegl, 2003; Hafed & Clark, 2002). The general patterns of interactions of microsaccade rates and orientations is more complex (Engbert, 2006), however, a computational model based on the SAW model has been proposed (Engbert, 2012).

Early accounts of fixational eye movements conceptualized them as the result of random firing from oculomotor units (Eizenman et al., 1985), or else as a nuisance component that causes blurring, if not corrected by the visual system (Burak et al., 2010; Packer & Williams, 1992). More recent evidence points to fixational eye movement being a necessary and useful component of visual exploration in counteracting receptor adaptation (Martinez-Conde et al., 2004; Rucci & Victor, 2015). First, fixational eye movements prevent visual fading (Ditchburn et al., 1959; Martinez-Conde et al., 2006) caused by neural adaptation (Coppola & Purves, 1996; Martinez-Conde et al., 2004; Martinez-Conde et al., 2006). Although fading prevention may be achieved by drift alone, microsaccades are much more effective at restoring vision after fading has set in (McCamy et al., 2014; McCamy et al., 2012). Second, both drift (Boi et al., 2017; Rucci & Victor, 2015) and microsaccades (Poletti et al., 2013) have been found to facilitate high acuity pattern vision (Intoy & Rucci, 2020). Specifically, the performance of an edge detection model can be improved by the addition of a movement component (Schmittwilken & Maertens, 2022). In another study, A. G. Anderson et al. (2020) use a Bayesian model of neurons during early visual processing that simultaneously estimates eye motion and object shape. The authors also find drift motions benefit high acuity vision, mainly by averaging over the inhomogeneities in the retinal receptors and receptor density. Finally, microsaccades and drift have been found to be both corrective, i.e., moving the eyes back to the intended fixation position, and exploratory or error-producing, i.e., moving new details into the center of the visual field (Engbert & Kliegl, 2004). Microsaccades are typically preceded by a reduction in drift (Engbert & Mergenthaler, 2006; Sinn & Engbert, 2016). Thus, both microsaccades and drift are functionally related and interdependent. It is a research goal of the current study to analyze the relationship between slow fixational eye movement components and microsaccades based on quantitative modeling while taking into account individual differences observed in experimental data.

The SAW model integrates several of the above properties of fixational eye movement and is biologically plausible (Engbert, 2012; Engbert et al., 2011; Herrmann et al., 2017). The model describes physiological drift by assuming a self-avoiding (random) walk (SAW) confined in a movement potential that limits the movements to reproduce visual fixation. A random walk on a lattice represents the trajectory of the eye. As the random walk traverses the lattice, visited locations are activated (see Figure 4.1). The activation represents the memory process that keeps track of the recently visited locations. This generative model successfully reproduces both the persistent and anti-persistent statistical properties of ocular drift Engbert et al., 2011. An extension of the model (Herrmann et al., 2017) implements neurophysiological delays and thereby matches the characteristic oscillations found in the displacement autocorrelation. Within the SAW model framework, Engbert et al., 2011 proposed a

mechanism for generating microsaccades based on the activation in the SAW model. However, the model has been a qualitative account so far, as it was not attempted to quantitatively reproduce experimental data from human observers.

In the present work we implement the SAW model in a likelihood-based framework in order to enable Bayesian parameter inference from fixational eye-movement data. We estimate model parameters for individual observers and find that the estimated parameters represent individually characteristic spatial statistics of physiological drift. Based on the quantitative agreement between simulated and experimental drift movements, we explore the relation to microsaccades. We use the model's latent activation to investigate potential mechanisms for triggering microsaccades. Since fixational eye movements have a strong impact on the spatiotemporal input that the visual system processes, reproducing these movements from a computer-implemented mathematical model is essential for a better understanding of visual functioning.

4.2 Results

In a first step we define the SAW model and describe the computation of the model's likelihood function. In a second step, we use the likelihood computation to estimate parameters for individual human participants. Finally, we use the model to generate data and conduct posterior predictive checks and exploratory analyses concerning the relationship of drift and microsaccades.

4.2.1 The model

As a theoretical starting point, the SAW model generates a random walk which is statistically self-avoiding (Freund & Grassberger, 1992). The self-avoiding walk is implemented on an $L \times L$ lattice where nodes are given by (i, j) with $i, j = 1, 2, 3, ..., L$. Each node carries some activation $a_t(i, j)$ at time t, which can be interpreted as neural firing rates. Initial activation values are set to $a_{t=0}(i, j) = 10^{-1}$ for all (i, j). At time $t = 1, 2, 3, \ldots$, first, the current activation of each node (i, j) across the field decays, according to

$$a_{t+1}(i, j) = \epsilon \cdot a_t(i, j) \ , \tag{4.1}$$

where $\epsilon = 1 - (10^{\gamma})$, representing the speed of the process memory decay. Second, activation is added to the nodes along the walker's trajectory, i.e.,

$$a_{t+1}(i^\star, j^\star) = a_t(i^\star, j^\star) + 1 \ . \tag{4.2}$$

Next, we implement a rule for activating lattice positions $(i^\star, j^\star)$ along the trajectory. We define an ellipse which is drawn such that the positions at times t and $t+1$ are the foci of the ellipse. The parameter ρ represents the size of the minor axis of the ellipse; the numerical value is set to 12 units (see Fig. 4.1). The lattice positions $\vec{v}^\star = (i^\star, j^\star)$ are defined as all lattice sites within the ellipse, which are activated according to Eq. (4.2).

The discretized map $\{a_t(i,j)\}$ of neural activations can be interpreted biologically, since grid cells have been found in entorhinal cortex which keep track of previously visited locations (Killian et al., 2012).

In principle, the self-avoiding walk can produce persistent motion if parameters are selected appropriately. However, during visual fixation, human observers are able to keep the eyes at an intended target. Therefore, the model implements a movement potential, which confines the random walk and represents a mechanism of fixation control. As a consequence, the model can produce anti-persistent motion on the longer time scale. Thus, the self-avoiding motion in a potential could maintain fixation at the intended location despite the necessity of refreshing the retinal input. The (time-independent) confining potential u is centered in the lattice and takes the form

$$u(i,j) = \lambda \frac{\left(\sqrt{(i-\frac{L}{2})^2 + (j-\frac{L}{2})^2}\right)^{\nu}}{L^{\nu}} \,. \tag{4.3}$$

Within this potential, motion is controlled by the sum of the self-generated activation $a_t(i,j)$ and the potential $u(i,j)$, i.e.,

$$q_t(i,j) = a_t(i,j) + u(i,j) \tag{4.4}$$

In order to chose the next step of the walker, we compute the probability of the next eye position as π_n, consisting of $q_t(ij)$, self generated activation and potential, and a time-independent stepping distribution, which controls the size of the movements, i.e.,

$$\pi_t(i,j) = q_t(i,j)^{-\eta} \exp\left(-\left[\left(\frac{i}{r_i}\right)^{\phi} + \left(\frac{j}{r_j}\right)^{\phi}\right]\right) \,. \tag{4.5}$$

From $\pi_t(i,j)$ we select the eye position at time $t+1$ using a linear selection algorithm.

As a result, in each time step follows the sequence of, first, relaxation of the current activation, second, selection of the next eye position under consideration of a stepping distribution, and, third, increase of the activation values along the current trajectory. Our model's behavior is controlled by a number of free parameters. In this study we selected a subset of parameters for estimation, namely γ the speed of the relaxation, the size of the stepping distribution r_i and r_j, the slope of the stepping distribution ϕ and the slope of the potential λ (i.e. $\theta = [\gamma, r_i, r_j, \phi, \lambda]$). We set $\rho = 12$, $\nu = 3$ and $\eta = 1$, in order to constrain the model and to obtain numerically stable behavior during parameter estimation. The parameters selected for estimation are of primary interest, as their values are interpretable quantities that may give insight into the biological plausibility of the model.

4.2.2 Likelihood function: sequential computation

The observation that makes the model compatible with a likelihood-based approach is the fact that the likelihood L at each time for each location on the lattice is given by

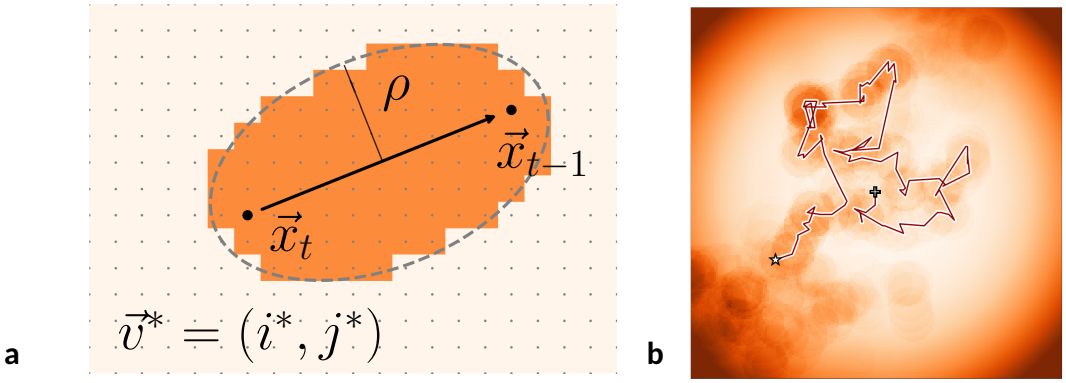

Figure 4.1 Illustration of the model activation. **a** Activation of lattice points within distance ρ from the eye's trajectory. **b** Simulated fixational eye movement trajectories, illustrating the SAW model. Starting from the position marked by the + and ending at the position marked by the star, the model generates activation along the trajectory. The movement is constrained by an activation potential centered around the fixation position. The activation profile is initialized by simulating some movement causing activation to be visible at locations that have not been visited. Activation decays gradually over time.

the selection map π. In other words given the time-ordered data $X_t = \{x_1, x_2 \ldots, x_t\}$, we can calculate the probability of observing the walker in position x at time t given the model and given all previous positions X_{t-1}, i.e. $P_{\mathrm{M}}(x_t \mid X_{t-1}, \theta)$. The likelihood of a sequence of t events is therefore given by the product of t conditional probabilities $\mathcal{L}_M(\theta|X_t)$, which is given by

$$\mathcal{L}(\theta \mid X_n) = P_{\mathrm{M}}(x_1 \mid \theta) \prod_{i=2}^{n} P_{\mathrm{M}}(x_t \mid X_{t-1}, \theta) \ . \tag{4.6}$$

In order to estimate the parameters of the model, we use Bayes' theorem to compute the probability of the parameters $\theta = [\gamma, r_i, r_j, \phi, \lambda]$ given the data as

$$P(\theta \mid X_n) = \frac{L_{\mathrm{M}}(\theta \mid X_n) P(\theta)}{\int_\Theta L_{\mathrm{M}}(\theta \mid X_n) P(\theta) \mathrm{d}\theta} \ . \tag{4.7}$$

A large literature exists to solve likelihood-based parameter estimation computationally. In order to leverage the full power of the likelihood-based approach we estimate the full Bayesian posteriors for each parameter using a differential evolution adaptive metropolis sampling algorithm (DREAM) (Laloy & Vrugt, 2012; Shockley et al., 2018). More details are provided in the Methods section.

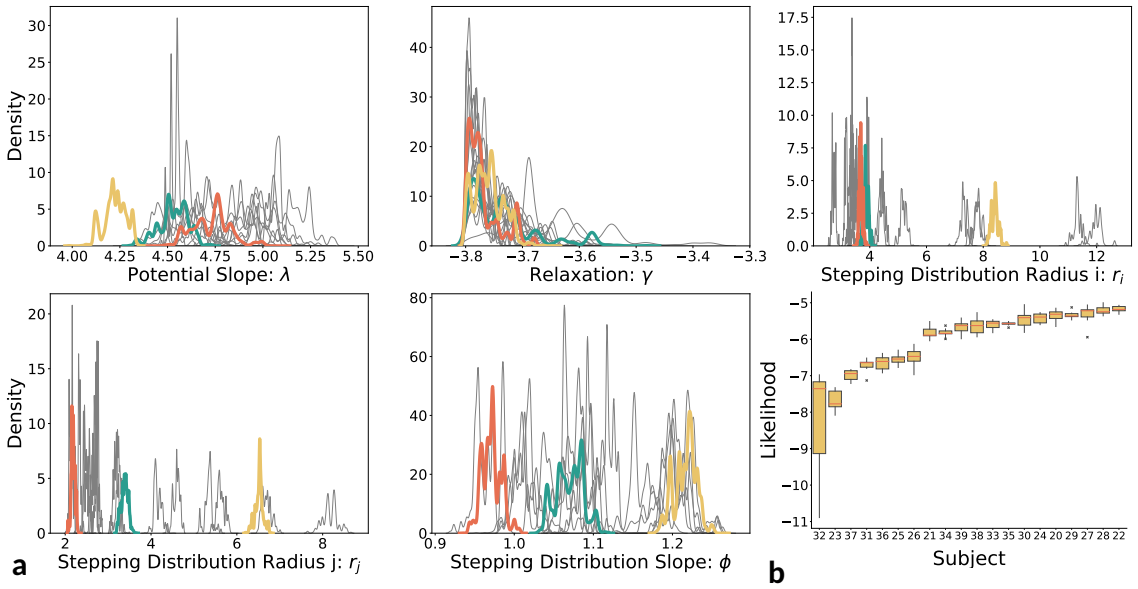

Figure 4.2 Parameter inference results.(a) Marginal posteriors for the five estimated parameters; the grey lines represent different participants, colored lines highlight three participants for comparison. For all five parameters, the posteriors converged to distinct peaks for the different participants. **(b)** A box plot based on the sum log likelihood of all trials, sorted by mean log likelihood across participants.

4.2.3 Parameter estimation results

We estimated the values of the free parameters of the SAW model for each subject independently based on data. The chosen priors (see Appendix for details) were relatively uninformative truncated Gaussians. Figure 4.2 presents the marginal posteriors for each participant in the study. In the Appendix (Table B2) we present the point estimates and 98% confidence intervals (Kruschke, 2014) for each parameter.

Biologically plausible models are designed to be grounded in real-world biological mechanisms and processes. As such, the values of their parameters reflect the known properties of these mechanisms. It is therefore informative to investigate the parameter values themselves, as they permit inferences about concrete aspects of the data generating process. First, the parameters r_i and r_j correspond to the widths of the stepping distribution. The final selection map in the model (Eq. 4.5) and consequently the resulting predictions for step sizes depend on the stepping distribution containing r_i and r_j, as well as the confining potential and activation. We define the step size as the distance travelled from one measured experimental sample to the next, i.e., 2 ms as we are using a sampling rate of 500 Hz. As shown in Figure 4.3C the parameters r_i and r_j are strongly correlated with the empirical step size distribution.

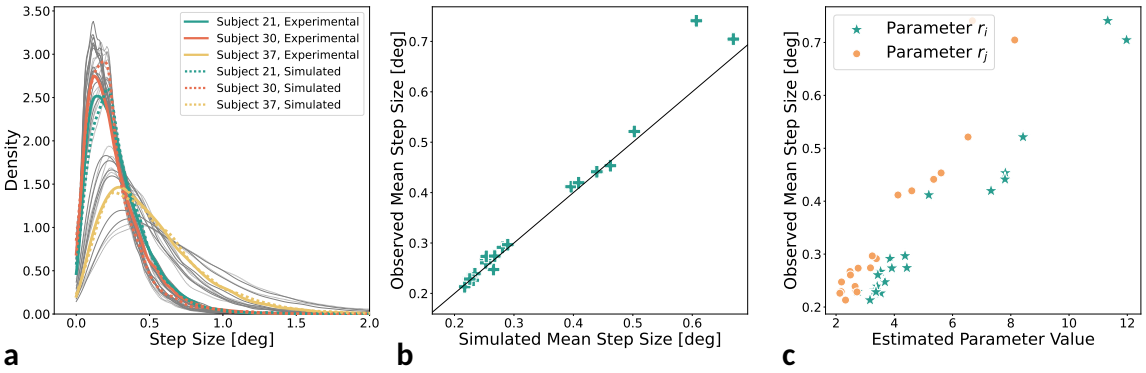

Figure 4.3 The fit between step sizes in the model and in the data.(a) shows the step size distribution, where highlighted lines represent individual subjects. **(b)** shows the correlation between simulated and empirical step sizes. **(c)** shows the correlation between the observed step size and the parameter value of r_i and r_j. As expected, the two parameters correspond directly to the stepping distribution in the data.

The parameter ϕ represents the slope of the stepping distribution. Higher values are associated with a stronger tendency to move along the cardinal directions. The estimated values for ϕ range from 0.9 to 1.3, indicating that the preference for cardinal directions is stronger in some participants than in others. The model captures individual differences in the stepping distributions, as demonstrated by the distinct posteriors obtained for ϕ, r_i and r_j.

Next, the parameter γ relates to the speed of memory decay in the process, where smaller, more strongly negative values cause slower decay, i.e., longer memory of the visited locations, and larger values closer to 0 cause faster memory decay, i.e., shorter memory. It was bounded in the estimation to a minimum of -4, as smaller value, i.e. less relaxation, caused the activation in the system to continuously increase, making the model numerically unstable. We find an average value of γ of 3.75 which indicates that activation decreases by 25% over the course of a trial duration of 3 s. Parameter γ accounts for relatively small individual variability. For most of the participants, estimates indicate a long memory for activation that represented the past trajectory in the system.

Finally, parameter λ represents the slope of the confining potential. The shape parameter of the potential function was fixed to 3, fixing the qualitative form to a steeper gradient than in a quadratic case (Engbert et al., 2011). The slope was considered a free parameter for the estimation. As the different participants' data lead to different decay and stepping parameters, the slope of the potential needs to accommodate different resulting values of mean system activation.

4.2.4 Posterior predictive checks

Using the estimated parameters (i.e., the posterior parameter distributions), we simulated artificial data sets of the same size as the experimental test data sets. We know that the model is, in principle, able to recreate statistical tendencies of ocular drift found in experimental data on a qualitative level (Engbert et al., 2011). Posterior predictive checks serve to show that the fitted model reproduces the expected statistical tendencies, i.e., turning angle- and step size distributions, as well as whether individual differences are reproduced.

Fitted primarily by parameters r_i and r_j, the step size distribution is represented in Figure 4.3 A. As defined above, the step size is the spatial distance between two subsequent samples. In order to condense the distribution of step sizes into one summary value, we computed the mean of each distribution. Thus, Figure 4.3 B shows the correspondence between the mean true and simulated step sizes. The model fits the step size very precisely and perfectly captures the differences in the individual preferred step size.

Next, we investigated the absolute and relative turning angles. There exists a preference in both ocular drift and (micro)saccades for movements in the cardinal directions, which may be caused by the structure of the ocular muscles (Sparks, 2002). This fact is captured well by the model (Figure 4.4A). The relevant parameter responsible for this effect is the stepping distribution slope ϕ, which shapes the stepping distribution to have a stronger or weaker preference for the cardinal directions. In order to ascertain whether the differences in individual behavior can also be reproduced, we use the area under the cumulative density function as a summary statistic (see the Methods section). The result in Figure 4.4B shows that the model is capturing individual differences. Furthermore, there is also a tendency for movements to be in line with or orthogonal to the previous movement vector. While this tendency is a lot more pronounced in the experimental data than in the simulated data, the simulations do show a qualitative reproduction of the trend (Figure 4.4C). However, it is likely that the peaks are driven by the preference in absolute angles, as no mechanism for relative angles was built into the model. We propose that the difference between the simulated and experimental relative turning angle distributions reveals the part of the relative turning angle distribution that must be accounted for by an additional, independent model mechanism.

Empirical ocular drift data, has been found to be persistent at small time lags and antipersistent at longer timescales. As shown in Figure 4.5A, the persistent trend in our data is not as pronounced for all participants as in comparable previous studies (Engbert & Kliegl, 2004; Herrmann et al., 2017). The antipersistent period begins around 60 to 80 ms after stimulus onset. The data simulated using the fitted parameters reflects this behavioral change well. However, it behaves more randomly than truly antipersistent at short timescales and becomes too strongly persistent at long timescales. Experience with the model shows that it can produce truly antipersistent behavior by varying the free parameters. Possible reasons why this trend was insufficiently captured by the fitted models are that the MSD is a less dominant tendency

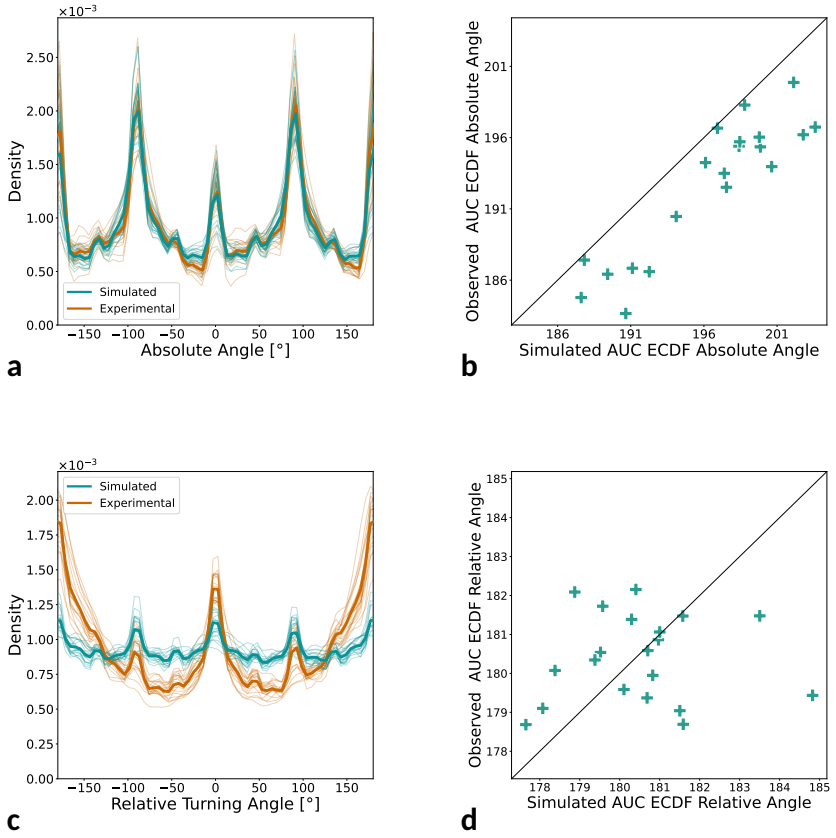

Figure 4.4 Turning angle distributions. (a) and **(c)** show the absolute and relative turning angles of one sample to the next. Thin lines represent individual subjects and thicker lines represent the means. Simulated data is shown in green while experimental data is red. Panels **(b)** and **(d)** show the respective correlations between simulated and experimental data. We characterize the angle distributions using the area under curve (AUC) of the empirical cumulative density function (ECDF). For details see the Methods section.

compared to other statistics and that our selection of free model parameters were correlated in a way that limited the ability to fit this particular tendency. Alternatively, the self-avoidance of the model may not be the only cause for early persistence, suggesting a model with an explicit exploration mechanism. Nonetheless the qualitative change in behavior is clearly present in the simulated data. Accordingly, the correlation between the experimental data and the simulated data show that the present model fits only account for a small part (roughly 10%) of the individual variation (4.5B and C).

4.2.5 Investigating microsaccades

Previous work concerning the SAW model has suggested a connection between the self-avoiding properties of the random walk and microsaccade triggers. A reduction in

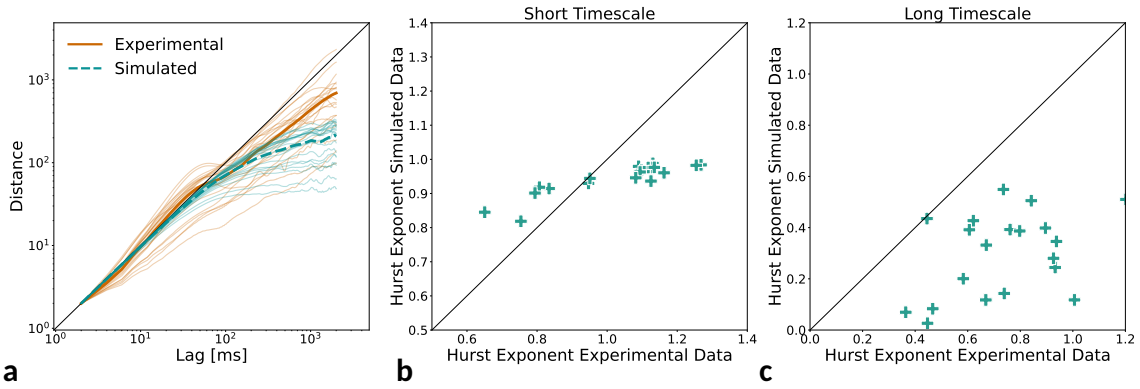

Figure 4.5 Persistent and Anti-persistent behavior. (a) shows the MSD of simulated and experimental data, where thinner lines represent individual subjects and thicker line represent the averages of experimental data in red and simulated data in green. Note that the lag is given in ms, whereas the distance is given in normalized units, in order to visualize the slopes in comparison to the identity line. **(b)** and **(c)** shows the correlation of simulated and experimental data of linear fits of the Hurst Exponents for the short and long timescale for individual subjects.

movement tends to precede microsaccades (Engbert & Mergenthaler, 2006). Following this reasoning, in the model a reduction in movement corresponds to a build-up of activation in the current position. Thus, we investigated whether the activation predicted by the model is indeed related to the occurrence of microsaccades. Specifically, we calculated the activation $q_t(i, j)$ at times relative to microsaccade onsets t_{MS-on} and offsets t_{MS-off} using the experimental data. Figure 4.6 shows the average activation around the time of a microsaccade, which is consistent with the hypothesis that high levels of activity are related to triggering microsaccades. In order to better understand the extent of the effect, we randomized the microsaccade onsets within each subject and computed the same trajectory on the randomized data. We find that the activation rises more before a microsaccade and drops more steeply compared to the randomized controls.

4.2.6 Model comparisons

The proposed model comprises three main components: the random walk with a stepping distribution, the self-activated trace memory, and the potential. In theory, the combination of both creates an interplay between persistence and fixation control. In order to better understand the role of each component, we created 3 control models by removing individual components. Specifically we investigate

1. the full baseline SAW model, that contains all three components
2. a model that is a random walk in a potential (W),
3. a model that is a random walk without a potential(W-NP),

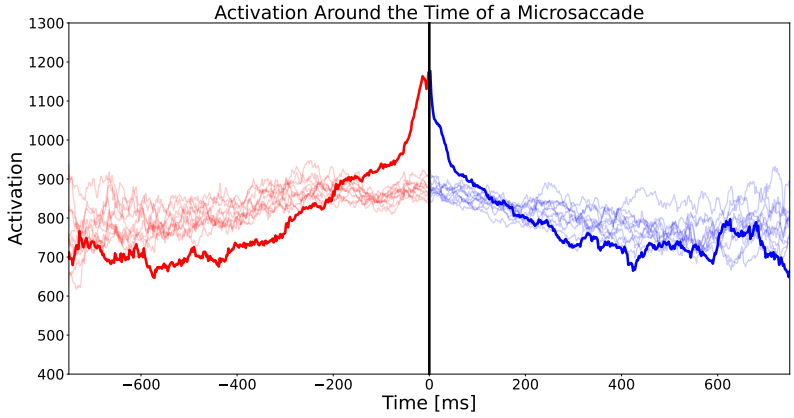

Figure 4.6 The average activation in the model around the time of a microsaccade. The less opaque lines represent randomized controls, where microsaccade onsets were randomized over the trials within one subject. In the data we find a distinct rise in activation before the time of a microsaccade and a drop in activation after.

4. a model that is a random walk with self-avoidance but without a potential (SAW-NP).

Note that the latter two models differ from the first two significantly in that they are not generative and, therefore, are not biologically plausible. In the absence of a potential it is still possible to compute the likelihood, but it is not possible to usefully simulate data from them, as there is nothing stopping the walker from simply walking away. We include them here, because they provide a relevant comparison, however, these models must be treated as substantially different concerning the conclusions they permit.

First, we compare the models in terms of their likelihood (Figure 4.7A). Each model was evaluated on the test data set, using the same parameters wherever applicable. Our findings show that in the generative models, SAW outperforms the version without self activation (W). Removing the potential and maintaining the self-avoidance (SAW-NP) also reduces performance. However, we also find that the non-generative model without either potential or self-avoidance (W-NP) shows almost identical likelihood to SAW. This suggests that individually fitted stepping distributions and general random walk behavior by themselves capture the most predominant features of the data. It is important to note that the likelihood is a very general measure of model performance-this model (W-NP) is neither biologically plausible, nor can it capture any additional statistical properties of the data. Thus, the lesson to be drawn from this finding is twofold: First, a high likelihood is not always a guarantee of an appropriate model and, second, that either a potential or self-avoidance by itself are not beneficial to model performance. Each component, added individually, actually reduces model likelihood. The fact that their joint effect reestablishes a similar likelihood while simultaneously making the model more biologically plausible should be considered a success.

To this comparison we add another model: the SAW model without individual

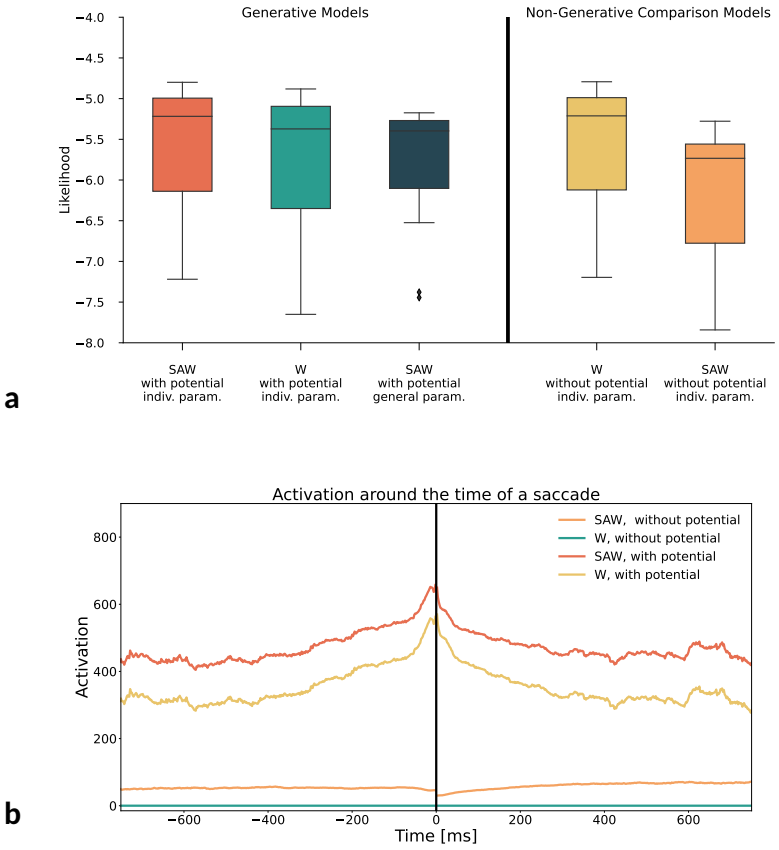

Figure 4.7 Comparisons of the different model versions. Panel **(a)** shows the 5 different versions of the model, including different components. We show that among biologically plausible models that the proposed mechanisms and the individual fitting procedure confer a likelihood benefit. Among Non generative models a pure random walk with fitted step sizes, while not biologically plausible, also achieves a high likelihood score. Panel **b** shows the model activation around the time of a saccade. The activation peak emerges only in the models that contain a confining potential.

parameter fits by subject. We observe that on average, there is a distinct benefit of fitting individual subjects. However, the averaged parameter model not only reduces high likelihood values but also the number of very low likelihood values.

Second, using our comparison models, we investigate which components drive the microsaccade effect. The very clear picture in Figure 4.7 shows that the activation peak is present only for the two models with a potential. This indicates that the effect is not driven, as we supposed, by a build-up of self-generated activation but rather by the potential. This is consistent with the idea that the microsaccades that drive the effect are related to control of the fixation position.

4.3 Discussion

Fixational eye movements display a large degree of randomness and both its origin and purpose have been much debated in the literature. Mathematical modeling approaches have contributed insights into the neurophysiological origin of the movement (Ben-Shushan et al., 2022; Eizenman et al., 1985), the desirable or undesirable consequences of the motion on image processing (A. G. Anderson et al., 2020; Schmittwilken & Maertens, 2022), and the spatiotemporal statistics of the drift trajectories (Burak et al., 2010; Engbert et al., 2011; Roberts et al., 2013). We performed Bayesian likelihood-based parameter inference of a self-avoiding random walk model for fixational drift at the level of individual observers. The estimation of the parameters converge to distinct marginal posteriors and data simulated on the basis of the fitted models reproduces individually characteristic behavior. In a second step we propose a relationship between the microsaccade rate and peaks in the model's latent activation state. This intuition is confirmed by an exploratory, data-driven analysis.

4.3.1 Individual variability

Fixational eye movements are controlled by a complex combination of factors, such as oculomotor control, attention, and cognition. The specific observed patterns vary greatly by individual both for measures of ocular drift (Cherici et al., 2012) and microsaccades (Poynter et al., 2013). Our results indicate that the individual variability in drift can at least partly be captured by the parameters of the SAW model. The average preferred step size is a particularly pertinent example (r_i, r_j), but also directional preferences (ϕ) and potential slope (λ) are different between subjects. These parametrizations are sufficient to simulate data which mirrors the characteristic features of individuals. To our knowledge this is the first paper that models individual variation of fixational eye movement trajectories.

The variability in drift between individuals has been found to be related to the individual variability in acuity (Clarke et al., 2021). Individual differences in fixational eye movement may therefore be related to a range of factors including the precise acuity of the eye, the tendency to maintain precise fixation (Cherici et al., 2012). Moreover, attentional preferences found in macroscopic eye movement during facial feature viewing translate to microscopic eye movement preferences (Shelchkova et al., 2019).

We find a distinct benefit from using individually fitted data sets over using a single averaged value for each parameter. This benefit becomes apparent in better convergence properties of the parameter estimation, as the strong distinct influences otherwise cause complex, multi-modal posteriors. Moreover, the individual fits cause a higher overall likelihood and fit of individual characteristics. However, while model based in averaged parameter values has a lower likelihood, it also reduces the number of very low likelihood data points. Thus, individuals who's data can not be easily fitted, would benefit from a normative influence of other subjects. This suggests the use of a hierarchical Bayesian modelling approach in future studies.

4.3.2 The confining potential

The characteristic Mean Square Displacement (MSD) of fixational drift is persistent at small time lags and anti-persistent at longer lags (Engbert & Kliegl, 2004; Herrmann et al., 2017). This is consistently the case, when averaging over large amounts of data. On an individual level, however, this tendency is not always equally pronounced. While the SAW model reproduces the transition to antipersistent behavior well, it does not adequately represent the persistent component, with the current data. As, in principle, self-avoiding random walk models can produce persistent behavior (Engbert et al., 2011; Roberts et al., 2013), this may be caused by a number of factors including the selection of fitted free model parameters, the relatively low persistence in the present data set, or a dominance of other statistical tendencies. Alternatively, it is possible that the strength of the persistent trend, which the SAW model frames as the result of self-avoidance, is in fact amplified by an additional explicit exploration mechanism which remains to be identified.

Exploration, or the explicit persistence, of the trajectory is highly variable, even between trials. The confining potential in the SAW model is static, representing a fixed intended fixation position. This is most likely a simplification, as the intention may change over time. Experimentally, we find a large amount of variation in the cohesion of the drift. In some trials drift consistently occurs around a specific position. In others it is evident that two intended fixation positions coincided over the course of the trial. In yet others, drift consistently maintains its direction away from the starting point. As the task and stimulus in the experiment was the same for all trials there is little evidence to explain this variation, aside from random variation or influence of recent past stimuli. A potential future direction for the model could be to implement a dynamic confining potential which is centered around a moving average of a number of recent samples. The limitation underscores the need for more comprehensive and accurate models to better capture the complex and individual characteristics of ocular drift behavior.

4.3.3 The relationship between drift and microsaccades

Early hypotheses suggested that microsaccades serve a corrective function for ocular drift (Cornsweet, 1956; Ditchburn & Ginsborg, 1952; Nachmias, 1959). Experimentally, however, no reliable correlation data confirms this hypothesis, as microsaccades can be explorative as well as corrective. Specifically at shorter time scales microsaccades induce persistent correlations while at longer time scales they tend to reverse movement to correct the fixation position (Engbert & Kliegl, 2004). Additionally, studies have shown that during high-acuity observational tasks, participants naturally suppress microsaccades without training (Bridgeman & Palca, 1980; Winterson & Collewijn, 1976), leading to the conclusion that microsaccades may serve no useful purpose (Kowler & Steinman, 1980). However, more recent research has demonstrated that microsaccades can enhance the visibility peripheral stimuli (Martinez-Conde et al., 2006), facilitate high acuity vision (Intoy & Rucci, 2020; Poletti et al., 2013) and

are responsive to task demands (Ko et al., 2010), suggesting a direct link between microsaccade activity and visual perception.

Thus, the role of microsaccades in fixational eye movement and their relationship with drift is not fully understood. Engbert and Mergenthaler (2006) suggested that microsaccades are triggered by a reduction in drift movement, i.e., low retinal image slip. This idea was further explored by the suggestion of a relationship between the self-avoiding random walk model and microsaccade triggers (Engbert et al., 2011). Our study provides further evidence supporting this connection. A build-up of activation in the current position of the SAW model, is associated with the occurrence of microsaccades. This finding is consistent with previous work indicating that a decrease in movement precedes microsaccades (Engbert & Mergenthaler, 2006). Our data-driven analysis revealed that the activation rises more before a microsaccade and drops more steeply compared to the randomized controls, which indicates that high levels of activation are more likely to trigger microsaccades. By comparing model variations we find that this trend is primarily related to the potential, indicating that the portion of microsaccades we capture with our analysis are related to fixation control. However, the exact mechanism behind this relationship and the potential causal direction between activation and microsaccades requires further investigation.

4.3.4 Other trajectory models

Physiological drift, as the slow component of fixational eye movement, is often modeled as a random walk (Burak et al., 2010; Kuang et al., 2012). Particularly when it is considered mainly as a component in a model of visual processing (e.g., Schmittwilken & Maertens, 2022), this approximation can yield good results. However, the statistical properties of the trajectories do differ significantly from simple randomness. To better capture these aspects a self-avoiding random walk has been proposed (Engbert et al., 2011; Roberts et al., 2013). However the number of models that aim to predict fixational movement trajectories is very limited. The SAW model used in this paper is one example. Another self-avoiding random walk model was published by (Roberts et al., 2013). Instead of an elliptical activation trace Roberts et al. (2013) implement the self-avoidance by choosing each step direction from a continuous distribution that is weighted by the density of recent gaze history in each direction. It achieves a similar result: at short time scales the model is persistent, avoiding previously visited areas. The memory of the process is limited and parametrized, allowing the authors insight into the process memory by estimating parameters. Due to the lack of an constraining potential, this model does not represent the subdiffusive component at long time scales.

4.3.5 Input dependence

Although initially fixational drift was often characterized as noise produced by the oculomotor units (Eizenman et al., 1985), more recent evidence from electrophysiological recordings in monkeys shows that fixational drift originates higher up in the

chain of command than the oculomotor neurons (Ben-Shushan et al., 2022) and is influenced by attentional processes (Shelchkova et al., 2019). Microsaccades, too, are influenced by attention and preferentially move in the direction of the attended region when there is covert attention (Engbert & Kliegl, 2003; Hafed & Clark, 2002). Thus, drift and microsaccades depend on task and the features of the fixated target (Bowers et al., 2021). This interplay of perception and action is consistent with the idea of active vision (Findlay & Gilchrist, 2003), even at the scale of fixational eye movement.

The SAW model is stimulus independent. Other modeling approaches have investigated the interdependence of fixational eye movements and visual perception. It can be shown that the visual processing stream is quite capable of dealing with the hypothesized motion blur caused by the constant displacement of the stimulus over the receptors (Packer & Williams, 1992). In fact, fixational drift is beneficial for high acuity vision, presumably because it allows spatial information to be redistributed into the temporal domain, modulating the input to individual receptors (Clarke et al., 2021). Image-computable models of edge detection can actually be improved by introducing drift (Schmittwilken & Maertens, 2022).

Research investigating the role of motion in visual perception typically use a random walk and are most likely quite robust to changes in the precise type of motion. However, the experimentally observed statistical properties of drift differ from simple random noise. More research is needed to ascertain whether these properties convey additional benefits to visual processing. In a recent paper A. G. Anderson et al. (2020) suggested a joint approach to infer movement and stimulus simultaneously. The model assumes a grid of retinal cells onto which stimulus patterns are projected and that the visual processing system does not have access to an efference copy of the movement. Instead they use Bayesian inference to alternatingly estimate the movement and the stimulus from the spike rate generated by the retinal cells. The authors conclude that drift is beneficial for high acuity vision as it helps the system to average over inhomogeneities in the retinal receptors.

Thus, fixational eye movements play an important role for visual processing. Integrating the fact that it is both stimulus-dependent and individually characteristic, suggests that the movement is optimized to account for individual physiological differences. This is consistent with the finding that fixational eye movement and visual acuity are related (Clarke et al., 2021). A future direction for fixational drift research may be to implement the idea that drift improves visual acuity in a generative model, to infer the ideal motion to prevent fading or to enable edge detection. Furthermore, although visual processing has been found to be quite robust, the development of more accurate models of fixational eye movement may improve the quality of models of visual processing.

4.3.6 Conclusion

In conclusion, our study contributed to fixational eye movement research through the application of mathematical modelling and Bayesian likelihood-based parameter inference. Our analyses suggest that self-avoiding random walk models can effectively

capture individual fixational drift behavior, as evidenced by the convergence of distinct marginal posteriors for each observer. Furthermore, our data-driven analysis indicates a relationship between microsaccade rate and peaks in the model's latent activation state, providing further insight into the underlying neurophysiological mechanisms of microsaccade triggering. Overall, our results provide a valuable contribution to the understanding of fixational eye movements and highlight the importance of individual differences in this behavior.

4.4 Methods

4.4.1 The likelihood-based modelling framework

Biologically motivated, mechanistic models allow researchers to test whether the proposed mechanisms are capable of producing the observed behavior, to identify which components are essential, and to explore how changes to the system's structure alter its output (Bechtel & Abrahamsen, 2010). Historically, the standard approach for cognitive models involves comparing them to time-independent summary statistics, e.g., here it may be MSD. The likelihood-based approach offers a number of advantages. First, it is possible to estimate the model parameters from the data in a fully Bayesian and statistically rigorous way. The value of the model is independent of any particular ad-hoc metric, the researcher may want to investigate (Schütt et al., 2017). The model likelihood can further be used as a basis for model comparison. Lastly, using the estimated parameters to simulate data, it is possible to conduct posterior predictive checks using metrics such as MSD to investigate whether the data constrains the model in a way that produces the expected behavior (Engbert, 2021). Thus, likelihood-based parameter inference allows compelling conclusions about the underlying mechanisms. Another advantage is that Bayesian parameter estimation provides a natural way to quantify uncertainty in the estimates, through the use of posterior distributions and credible intervals. This can be especially useful in cases where the data are noisy or the model is complex.

By independently estimating separate parameters for each experimental subject, it is possible to investigate individual differences. The parameter estimation yields a separate posterior distribution for each subject. As the parameters represent interpretable quantities with biological counterparts, the comparisons of the posteriors can allow interesting insights. Additionally, when a model is capable of representing individual differences, it speaks to the validity of the model and its parametrization.

4.4.2 Experimental data

The experimental data used for this study were eye movement trajectories recorded using an Eyelink 2 with a sampling rate of 500 Hz. Participants were seated at a distance of 50 cm to the monitor and calibrated using a 9-point calibration grid. Each trial consisted of a fixation task, where participants fixated a cross in the center of a

white screen for 3 seconds, followed by a scene viewing task. Here, we use only the enforced fixation data from the first 3 seconds. Out of 50 recruited participants, 48 completed all 40 trials. A further 6 were later excluded due to a large number of blinks. As the experiment included a rigorous online quality control and the calibrate-ability of subjects varies, 2 participants aborted the experiment. Trials during which saccades or blinks were detected were repeated immediately. In order to detect microsaccades we used a velocity-based algorithm (Engbert, Sinn, et al., 2015). The data set is publicly available on the Open Science Framework (www.osf.org/fbuxq)

4.4.3 Parameter estimation

Here, we used the DREAM$_{ZS}$ algorithm (Laloy & Vrugt, 2012). Rooted in the classical Metropolis (Hastings) Markov Chain Monte Carlo (MCMC) algorithm, DREAM$_{ZS}$ iteratively explores the parameter space by sampling its position and asymptotically converges to the true posterior distribution of the parameters. The algorithm includes several (Markov) chains starting at arbitrary (random) positions in the parameter space. For each chain new positions are chosen by combining (hence "evolution") positions of other randomly chosen chains, including their past states.

In Bayesian parameter estimation, the unknown parameters are treated as random variables and are assigned a prior probability distribution. This prior distribution reflects the researcher's initial beliefs about the likely values of the parameters based on prior knowledge or experience. We chose relatively broad truncated Gaussian priors, which did not constrain the estimation very strongly. The truncated tails were chosen according to experience with the model to prevent numerical problems in the case of extreme parameter values.

We split this data set into two separate sets: one half (20 subjects) was used for model development and exploratory analyses. The other half (21 subjects) was used for the final analyses and model evaluation. Each set contains data from 27 trials for each subject. We discarded all trials where the movement during fixation exceeded 1.2 degrees of visual angle. Due to the high individual variability in the data this criterion excluded 7 subjects, because too many trials were affected. The 27 trials can be split into training and test sets, with a 2/3 to 1/3 split. Each trial was 1500 samples long, i.e. represented a fixation of 3 seconds. This procedure finally yielded a data set with equal numbers of samples per trial and trials per subject, facilitating statistical analyses.

4.4.4 Angle distribution comparisons

In order to compare and correlate the angle distributions of individual participants as well as between simulated and observed data, it was necessary to reduce the angle distribution to a single summary value. An area under to curve (AUC) metric is a common solution to this problem. However, in our case the distributions were already densities, i.e. their AUC was, by definition, equal to 1. Instead we compare the AUC of the empirical cumulative density function (ECDF). The ECDF is obtained by

sorting the observations into unique bins and calculating the cumulative probability for each. By grouping and computing the ECDF AUC of each group, we can compare the similarity in the peak height of the angle distributions.

4.5 Acknowledgments

This work was supported by Deutsche Forschungsgemeinschaft (DFG) via Collaborative Research Center SFB 1294, Project-ID 318763901. We thank Andrea Miroslava Pomar Robles and Stefan Seelig, who collected the experimental data set which was used in this study.

5 Discussion

With four parameters I can fit an elephant, and
with five I can make him wiggle his trunk.

John von Neumann

Eye movement and visual perception are complex and dynamic processes that play a critical role in how we perceive and interact with the world. A vast body of experimental and theoretical research has advanced our knowledge of visual processing. Situated within this rich research tradition, the dynamical approach to modeling cognition adopted in this thesis is consistent with the observation that perception and action are interdependent and unfold dynamically over time. Moreover, the presented models are mechanistic in the sense that they rely on biologically plausible mechanisms to generate behavior. The modeling framework I describe and apply here builds upon the extensive literature on dynamical modeling and Bayesian likelihood-based parameter inference. In the previous chapters I presented two models of human eye movement within this framework: the SceneWalk model for modeling macroscopic scan paths and the SAW model for generating fixational eye movements. Both models were fitted to experimental data independently for each individual subject (and task, in the case of SceneWalk in Chapter 3) using Bayesian parameter inference. Both models allow an examination of the temporal dynamics of their respective processes and were shown to capture important aspects of behavior as well as individual differences. In the following sections I will discuss the models specifically within the context of their respective literature and consider possible future research directions. I will conclude by commenting on the general significance of the presented approach and its methodological advantages.

5.1 Insights from modeling scan paths

Eye movements are a necessary component of visual perception. As the eyes scan an image, different areas move into the high acuity fovea. The resulting sequence of fixations and saccades is closely linked to the concept of visual attention (e.g., Schneider & Deubel, 1995) and provides a window into how visual processing functions in the brain. Broadly, the selection of each subsequent fixation is guided by a range of

attentional processes (e.g., Tatler, 2007), bottom-up (e.g., Itti et al., 1998; Mannan et al., 1996), and top-down (e.g., Henderson, 2003) influences. The following paragraphs discuss the SceneWalk model in the context of these guidance principles.

5.1.1 Distribution of attention

The distribution of attention over time guides fixation selection. Examples of this are turning angle- and saccade length distributions (Tatler et al., 2017), inhibition of return (Klein & MacInnes, 1999; Mirpour et al., 2019), and the central fixation bias (Rothkegel et al., 2016, 2017). The SceneWalk model, at its core, is an implementation of hypotheses for mechanisms that may cause these systematic statistics such as saccadic momentum (Rothkegel, Schütt, et al., 2019; T. J. Smith & Henderson, 2009), foveation (Parkhurst et al., 2002), an inhibition stream with distinct temporal dynamics (Klein, 2000; Klein & MacInnes, 1999), and transient strategical preference for the center (Rothkegel et al., 2017), respectively. The fact that the model does, indeed, produce the expected statistical properties in simulated data (as demonstrated in posterior predictive checks), adds evidence for the implemented mechanisms.

One such mechanism that we integrated into the SceneWalk model are pre- and post-saccadic attentional shifts. Experimental work shows a tendency toward improved accuracy and speed when reporting on stimuli at the upcoming fixation location just before a saccade is executed (Deubel & Schneider, 1996; Irwin & Gordon, 1998; Rolfs et al., 2011) and that a similar improvement can be observed at the same retinotopic location after the saccade is executed (Golomb et al., 2008; Marino & Mazer, 2016). The pre-allocation of attention in particular is a necessary component of saccade preparation. Its effect can be measured even when attention is directed away (Castet & Montagnini, 2007) and when the target is out of saccadic reach (Hanning et al., 2019). Our implementation of this concept in the SceneWalk model (Chapter 2) led to better model performance and, particularly a better fit of the turning angle distribution. The successful implementation of this mechanism provides evidence for the importance of attentional pre-allocation around the time of a saccade and allows us to explore the consequences for eye movement behavior. This result is particularly notable because it shows how attentional effects at the microscopic level, which are primarily subject of neurocognitive and psychophysics research, have a substantial effect on fixation selection. It highlights how very basic consequences of the physiological, functional architecture of the visual system, as well as movement kinematics are highly relevant to understanding higher-level decision-making processes.

It is important to note that the implemented attentional mechanisms, for example these specific pre- and post-saccadic shifts, as well as facilitation of return, were selected because of their pertinent and clear interpretation and implementability and do not, or indeed *can not*, represent a full description of *all* attentional mechanisms. There is a growing literature about additional attentional influences, which illuminates how visual attention is continually updated and redistributed in various dynamical ways. One study found that the sensitivity to features at the current fixation location is influenced by features at the upcoming fixation location (Kroell & Rolfs, 2022). Fur-

thermore, the sensitivity to specific spatial frequencies at the target location changes over time (Kroell & Rolfs, 2021) and the exact extent of the pre-allocated attention vary depending on grouping of features (Shurygina et al., 2021). These results illustrate how, in reality, the shifting attention profile includes a much higher degree of nuance and dynamics than is included in the SceneWalk model. However, the presented results support the claim that redistribution of attention around the time of a saccade is a central component of fixation selection.

The SceneWalk model seems to currently be unique in the fact that it combines a number of attentional mechanisms to generate a dynamical map of attention during scene viewing. Individual attentional mechanisms have also been implemented in models in the field of visual search (e.g., T. J. Smith & Henderson, 2009). In other scene viewing models similar systematic tendencies are used to evaluate the model, such as the turning angle distribution (Le Meur & Liu, 2015). However, in these cases the distribution of angles is explicitly added into the model, such that the fixation selection is in some way directly proportional to this experimental distribution. The implementation of a dynamically evolving attention profile is a key strength of the SceneWalk model, as it captures the complex interplay between attentional processes and oculomotor behavior that occurs during naturalistic scene viewing.

5.1.2 Spatiotemporal likelihood

Another newly added mechanism in the SceneWalk model is the interdependence of duration and location. The integration of fixation durations into the SceneWalk model is a natural extension of its dynamic architecture. Previously, fixation selection depended upon the duration via the evolution of the model over time. In the extended model the duration is generated using a rise-to-threshold model that depends on the saliency at the current fixation location. Previous research established there exists a connection between duration and selection (Tatler et al., 2017). The LATEST model, as discussed in Chapter 3, selects parameters using only fixation duration information and finds that the emerging spatial statistics show relevant characteristics. Similarly the WALD-EM model (Kucharsky et al., 2021) implements fixation durations as a information accumulation process and, like SceneWalk, uses a spatio-temporal likelihood approach to fit the model.

Both LATEST and WALD-EM fall into the category of information-accumulation models. In the SceneWalk model too, fixation durations are controlled by a continuous-time discrete-state random walk process (Laubrock et al., 2013; Nuthmann & Henderson, 2010). This process has the statistical property of generating fixation durations proportionally to a Gamma distribution. Thus, all three models agree that the information is accumulated over the course of the fixation and that fixation duration is therefore related to attentional processing. In our model implementation the accumulated information is represented as saliency at the current location. Ideally, instead of the saliency, we would use the model activation, as it unfolds dynamically over time. This would have been very computationally expensive, but would most likely correspond better to the biological constraints of the visual system.

Overall, the development of joint models of duration and location is an important step toward an integrated understanding of eye movement behavior. As duration is typically conceived as a measure of visual processing and attentional allocation, it is a natural complement to the location-based measures that have dominated the literature.

5.1.3 Bottom-up influences

Bottom-up, or image-dependent, aspects of fixation guidance are represented in the SceneWalk model by the underlying saliency map. This saliency map is explicitly not computed by the model, but instead required. In our work we usually used the experimental fixation density as a best-case estimate of saliency, but it is also possible to use model-generated saliency maps such as are generated by DeepGaze II (Kümmerer et al., 2017). As the SceneWalk model mainly focuses on scan path dynamics and has thus far not been fitted for individual images, its informative value concerning bottom-up guidance is limited.

It is relevant to note that the meaning of "bottom-up" and even "saliency", are used differently depending on the context. Early attempts at modeling considered exclusively low level features in order to represent an account of early visual processing (Itti et al., 1998). As the field has evolved, models have come to include a variety of high level factors (Kümmerer et al., 2017). This conveys an improvement for the prediction of fixations, but no longer maintains the aim of modeling low level visual processing. Particularly in the field of computer vision it has become common to refer to any distribution that maps features, relevance, meaningfulness ratings or conspicuity onto the 2D space of an image as a "saliency map". In fact, these ratings are likely to have some significant overlap, as features and meaning do tend to occur together. However, both experimentally determined fixation densities and deep neural network (DNN) models which are trained using experimental fixations, include both top-down and bottom-up information. Therefore, the SceneWalk model too, in its current implementation, bases its attention stream on an input that includes both bottom-up and top-down semantic and task information. A key advantage of the model formulation allows various saliency maps to be used. This opens the door to potential future research, concerning the impact of low-level versus high-level saliency on dynamical viewing parameters.

5.1.4 Top down: modeling individual differences

Top-down components of fixation selection refer to influences of observer differences, task instructions, and other high-level influences. In this thesis I present evidence that differences in the SceneWalk parameters can represent a tuning to both observers and tasks (see Section 5.1.5). In the domain of eye movement research it is well-established that a large part of the variation in the data is due to individual differences (Kliegl, 2010). There is a significant amount of psychological research indicating that macroscopic eye movements are unique to each individual. An extensive study involving

more than 1,000 participants demonstrated that these individual characteristics of eye movements are highly consistent and idiosyncrasies persist across multiple experimental sessions (Bargary et al., 2017).

We show that fitting models for individual observers yields stable and distinct parameters for each subject, even for relatively small data sets. This is a significant methodological improvement over models that omit parameter inference or rely on the increased size of grouped data sets. In fact, we found that parameter inference by subject has some unexpected advantages concerning the convergence. As subjects differ significantly in their viewing behavior, the reduction of variance actually improved convergence. However, the availability of sufficient appropriate data is a major bottleneck in modeling interindividual differences. This is particularly problematic in models with a large number of parameters, such as DNNs. It can also occur in other modeling approaches, when the fitting procedure is not efficient. This was the case in previous implementations of the SceneWalk model, which required fitting of parameters using ad-hoc performance metrics (Engbert, Trukenbrod, et al., 2015). Using a likelihood-based approach, as suggested by Schütt et al., 2017, made the parameter inference procedure more robust and statistically rigorous, as well as less data-hungry.

The ability of the SceneWalk model to converge to individually different posterior distributions of the model parameters suggests that the parameters at least partly capture the causes of individually characteristic behavior. This lends credibility to the mechanisms implemented in the model and is further supported by the fact that differences in parameters resulted in simulations that shared the individual characteristics of the behavior. Parameters relating to the fixation durations (t_β, q), to the saccade amplitude (σ) and to the degree of determinism in the target selection (γ) were particularly notable for their convergence to distinct peaks for individuals. This is consistent with the intuition that individuals differ in their processing requirements, i.e., attentional span or visual processing speed. The individual differences represented by the SceneWalk model allow for a more nuanced understanding of how visual attention operates in different individuals.

Eye movement differences between individual subjects are reliable and specific enough to be used for biometric identification (Jäger et al., 2020; Makowski et al., 2020). These approaches typically rely on DNN architectures and a large number of parameters to capture the differences. In a recent study, we (Makowski et al., 2020) used the SceneWalk model parameters as features in a discrimination model. This approach performed only slightly better than chance, although it was significantly improved by the addition of a Support Vector Machine (SVM) that uses a Fisher Kernel to transform the data into a more discriminative feature space. However, unsurprisingly, models that were specifically designed for the task of discriminating different viewers turned out to be considerably more performant. In general, it is likely that for the use case of biometric viewer identification fixational eye movement will turn out to be a more robust and less easily spoofed metric.

Other theory-based modeling approaches that include individual differences have also provided valuable insights into gaze behavior. Brodersen et al. (2008) explored the class of linear rise-to-threshold models of saccadic decision making on the basis

of three separate subjects in a learning task. The authors used likelihood-based parameter inference and found differences in the models between the subjects, revealing significant individual differences in the learning profiles and related eye movements. Coutrot et al. (2017) developed an Hidden Markov Model (HMM) to investigate task- and individual differences in gaze behavior. Overall, the dynamical modeling of individual viewing behavior represents a significant advancement of our understanding of gaze behavior. The presented research using the SceneWalk model fitted a comparatively large number of subjects and systematically investigated their individually specific behavior. Understanding between-subject variation is vital for a more comprehensive understanding of the underlying mechanisms, as it may provide an indication of the axes along which variance emerges.

5.1.5 Top down: task differences

In Chapter 3 we show that the parameters of the SceneWalk model for the same subject differ when different tasks are given. The observation that different tasks elicit different viewing behavior is well-established (DeAngelus & Pelz, 2009; Yarbus, 1967). Inferring task from eye movements has had mixed success- an early failure by both computational models and human observers (Greene et al., 2012) was later reanalysed and successfully classified using more advanced models Borji and Itti (2014). Later work by (Coutrot et al., 2017) used HMMs and found both individual and task differences. Although SceneWalk is not an application-oriented classification model, we also find differences in model parameters within an individual in response to different task instructions. These differences indicate that there is an attentional weighting and tuning to produce behavior that is best suited to the task demands.

The structure of the SceneWalk model implies a difference between dynamic aspects of scan path generation, implemented in the mechanisms of the model, and the static aspects, represented by the saliency information that is fed into the activation stream of the model. This is a simplification, since saliency does not exclusively represent bottom-up information. A more realistic conceptualization of the saliency information is that it also dynamically changes over time and is subject to changing top-down influences. Chapter 3 investigates the role of the underlying saliency, by using different fixation density maps as a basis: either general saliencies, averaged over task influences and therefore emphasizing bottom-up information, or task-specific saliencies, which are computed by task. As expected, the differences between tasks in the estimated parameter values are greater in the general saliency condition. As the general saliency does not explain any between-task differences on its own, the model parameters themselves must explain more variance. However, even when given task-specific fixation densities, the model produces different parameter estimates. This suggests that behavioral differences do not rely only on reweighting of certain aspects in the scene, but that the systematic components attentional spreading are adjusted in a significant way. Indeed, we found that task-specific parameter fits for dynamic components using general saliencies outperform a simple local saliency model with task-specific saliencies. These findings suggest that eye movements are not simply

determined by a reweighting of bottom-up saliency features, but also by a retuning of dynamical attentional mechanisms, that suits the demands of the task at hand. The SceneWalk model can serve as a valuable tool for understanding the dynamics of visual exploration and the interplay between attentional factors in a wide range of (naturalistic) settings.

Investigating the role of task is particularly relevant, as the use of the free viewing paradigm [6] has attracted considerable criticism. Aside from concerns about superficiality and lack of generalizability of laboratory setups like this, it has been suggested that when no task is given, participants, may, consciously or unconsciously, simply invent one (Tatler et al., 2011). As an alternative paradigm, one of the key strengths of visual search is that it is relatively easy to manipulate, making it useful for studying a wide range of research questions relating to attention control. The systematic nature of search also allows the identification of efficient versus inefficient approaches and has prompted theories regarding pop-out, conjunctive, serial, parallel search and more (see Wolfe, 2015, for a review). COCO-Search18 (Chen et al., 2020) is a notable and recent data set of goal-directed search fixations on natural scenes, which is intended to train and benchmark models for attention control in visual search. The authors present an evaluation of a number of competitive models of visual search on the data. Recent developments in eye-tracking technology, i.e., more mobile eye tracking setups, permit a new approach to the transferal of laboratory findings into the real world. It is clear that the interaction of perception and action extends beyond eye movements, and that real-world tasks require different viewing behavior than viewing static images (Matthis et al., 2018). On the other hand, many of the systematic findings from static scene viewing do translate into less restricted conditions (Backhaus et al., 2020). This area may be an promising future application of the SceneWalk model.

Overall, modeling both task differences and individual differences within the same cognitive mechanistic model is a powerful approach for understanding the mechanisms underlying attention control. The model demonstrates a considerable flexibility and precision by allowing both the specificity to different behaviors and in different observers. It may also provide relevant starting points for approaching the question of inferring task and observer from the data, which has the potential to have a wide range of applications, ranging from driving assistance and VR performance, to criminal investigations.

5.1.6 Contributions of other modeling approaches

The SceneWalk model aims to represent the underlying mechanisms that guide eye movement behavior. More generally, mechanistic models are built upon component cognitive, neural, and oculomotor assumptions and posit that the model's mechanisms correspond to actual processes used by humans during active vision. Other models of eye movement do not necessarily seek to match the mechanisms of eye movement

6 Typically, in free viewing experiments, participants look at static photographs of scenes, without any task instructions.

selection, but rather aim to account for statistical patterns observed in the data. Throughout this thesis I discussed models across the continuum from hypothesis-based to hypothesis free models. Notable examples that are discussed at length throughout the text are LATEST (Tatler et al., 2017), a model by Le Meur and Liu (2015), and WALD-EM (Kucharsky et al., 2021). Each has its own set of methods, strengths and weaknesses and provides insight from a distinct perspective. Here I would like to highlight three further alternative approaches.

First, the Exploration-Exploitation Model by Malem-Shinitski et al. (2020) applies the idea of the Exploration-Exploitation Dilemma, where a decision is made between exploiting information near the current gaze position and exploring other patches within the given scene. The model switches between internal states of local and global attention. In this model, we assume that the decision for each saccade is based on the available information. The probability of following the local attention map is higher if the ratio of priority values of the current and previously fixed locations is high. There is a dichotomy in modeling data, with one goal being to fit the data as precisely as possible and on the other hand mimicking the data generating process. The first is likely to make better predictions but for understanding the process deeply the second kind is of equal importance. In developing the SceneWalk model we started from the latter and added the inference procedure as it was necessary. The exploration-exploitation model of eye movement was designed particularly with the inference method in mind and thus takes the reverse approach. The exploration-exploitation model is a notable example of the value of interdisciplinary inspiration for advancing methodological research.

Second, in reality most eye movements do not take place on unmoving photographs of scenes. While the SceneWalk model is dynamic in the sense that the underlying process is driven by dynamical factors, the input stimulus remains static. In principle, in many cases, the same models could be applied to dynamic input. In practice, the challenges of eye movement prediction on video are considerable and include a high demand for computational resources. It also necessitates accounting for different types of eye movement, e.g., smooth pursuit, which does not occur in static scenes. One model that applies the principles of Bayesian inference is Zanca et al. (2020). Another example is a model by Roth et al. (2022), which implements many of the same mechanisms as SceneWalk, except that it is applied to video data. This highly promising extension of the mechanistic modeling approach is, at the present time still a work in progress, but represents a vital step in the direction of modeling every-day natural eye movement behavior.

Lastly, the DeepGaze III model (Kümmerer & Bethge, 2021) is situated firmly at the data-driven end of the hypothesis continuum. Recent advances in the use of deep neural network (DNN) models for predicting fixation sequences (see also Shao et al., 2017) have demonstrated that these models are capable of capturing a variety of scan path statistics and achieve high levels of prediction performance. DNN models rely on large data sets of eye movement data in order to optimize the many thousands of parameters that are the weights and biases of each component neuron. The quote at the beginning of this chapter puts this into perspective. John von Neumann's elephant

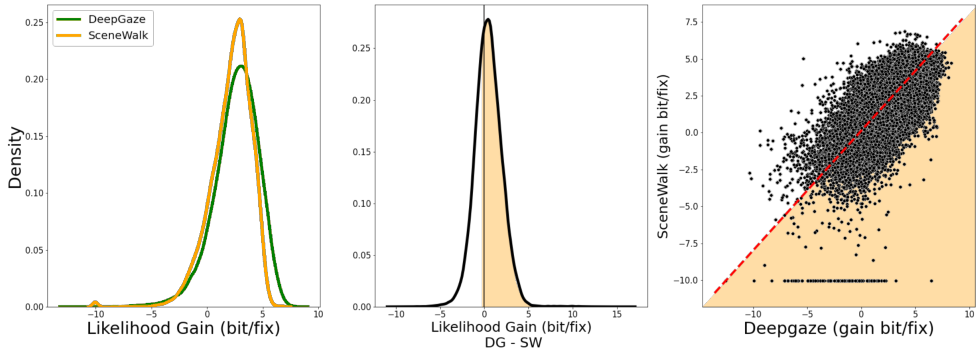

Figure 5.1 Fixation-level comparison between DeepGaze III and SceneWalk. The leftmost plot shows the density of likelihood values of both models. The more performant model, DeepGaze III displays a distinct shift to the right due to its higher predictiveness. The middle frame shows the resulting density when the likelihood gain of SceneWalk is subtracted from that of DeepGaze III for each event. The Third panel shows a correlation between the likelihood gain under SceneWalk and DeepGaze III. The areas marked in yellow highlight all the events for which DeepGaze III has a higher likelihood gain, i.e., predicts better.

represents a data set, which can be fitted more closely the more parameters are added. In the case of neural networks a similar principle applies—it is hardly surprising that SceneWalk, with its handful of free parameters, is outperformed by a DNN such as DeepGaze III. However, this does not diminish the value of the dynamical approach, as both approaches have their individual advantages and shortcomings. In fact, using both methods to complement each other has yielded fruitful results in an exploratory attempt, as described in the following section.

5.1.7 Contrasts and synergies of mechanistic models and neural networks

A model like SceneWalk is an implementation of known or assumed mechanisms. The behavior of the model is highly interpretable and it therefore offers a strong test of these hypothesized mechanisms. In order to fit the small number of free parameters, a comparatively small data set is sufficient. DeepGaze III on the other hand captures the data more closely but, on its own, provides little insight into the underlying processes. In a presentation for the Vision Science Society meeting 2022, I presented some research into how a synthesis of the two modeling approaches provide interesting insights into fixation selection. The presented idea posits that the DNN, with its superior performance, can serve as an effective estimate for how well the data can be predicted and how much of the variation in the data is individual or random variation. Meanwhile, the biologically-inspired models help to understand which mechanisms were learned by DNNs.

Both SceneWalk and DeepGaze III can be fitted using maximum likelihood estimation and both models compute likelihood predictions for each fixation. We investigated

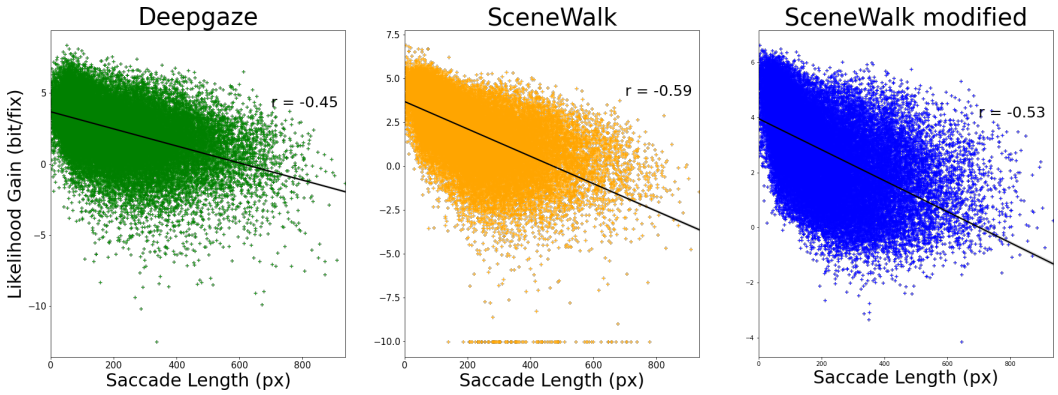

Figure 5.2 Correlation between saccade length and associated likelihood. The panels from left to right show the DeepGaze III model, SceneWalk, and the version of SceneWalk that was modified in the context of this study. Compared to DeepGaze III, SceneWalk has a stronger correlation, indicating less accurate prediction of long saccades. In the modified version this is slightly improved.

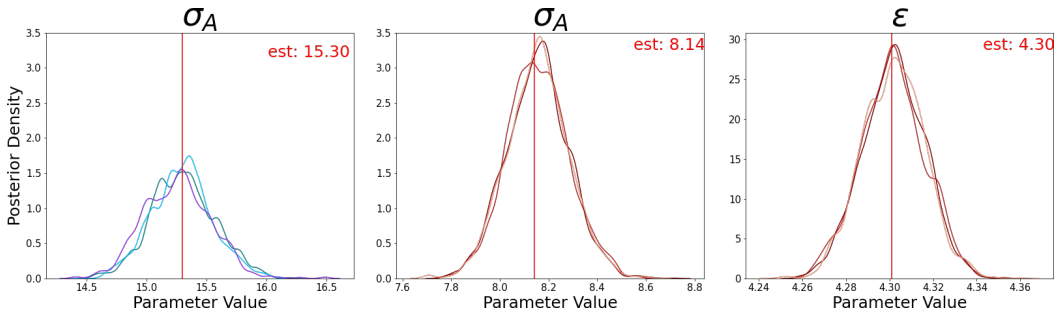

Figure 5.3 Marginal Posteriors for σ_A in blue on the left in the original model, and σ_A of the modified model in the center, and ϵ on the right. Both parameters of the modified model converge well, and their values enable a stronger concentrated attention in the center, while simultaneously permitting longer, exploratory saccades.

the fixations where the models diverge most in their prediction performance (Figure 5.1). This analysis revealed that long, explorative saccades in particular are often not predicted well by the SceneWalk model (see Figure 5.2). A qualitative case-by-case analysis reveals that these saccades occur often in images where multiple highly salient regions are at some distance to each other. The Gaussian window implemented as part of the SceneWalk attention stream effectively prohibits these long saliency-based saccades. Based on this observation we developed a modified SceneWalk model, using a function with heavier tails to represent the attentional window. Specifically, we added an additive parameter ϵ to the Gaussian Window G_A (Equation 2.1), effectively raising the distribution, before convolving the attention Gaussian with the saliency. We re-fitted the parameters and found that the modified parameters σ_A and ϵ converged to distinct posteriors. The tail-offset parameter ϵ was clearly non-zero, providing strong evidence for modification. The size parameter σ_A in the new model was smaller and had lower variance, indicating that that attention can be more localized when long saliency-dependent saccades are possible under the model (see Figure 5.3). The modified model had a slightly improved model likelihood but effectively reduced the divergence between the model predictions, specifically for long saccades.

This explorative research allows us to use the advantages of different types of modeling to advance the field of visual perception. The preliminary conclusion is that an integration of hypothesis-based and hypothesis-free modeling is a promising approach for understanding eye movement behavior. DeepGaze III captures a large part of the inter-observer consistency in the data through its large number of parameters. SceneWalk represents explicit knowledge; it explains only the variance which we have understood to a point where we can formulate it into a mechanistic model. The difference in model performance gives an indication of the remaining explainable variance (see Figure 5.4).

In order to explore this research direction further, we found a major hindrance to be the lack of appropriate data sets. Training DNNs requires large amounts of data, preferably with a large number of different images. SceneWalk on the other hand requires fewer data, as it has fewer parameters. It does, however, require long sequences, in order to be able to benefit from sequence effects. We found no open data sets that fulfilled the criteria of both models to an appropriate standard. In an ongoing effort we have preregistered an experiment and are currently collecting a large data set to serve this purpose: high data quality, long sequences and many images. More detail about the DAEMONS study may be found the Appendix E.

5.1.8 Future directions for the SceneWalk model

The SceneWalk model has been a useful tool for understanding attention control and search behavior in natural scenes. However, the potential for extension and modification remains large. There are several ways in which the model could be improved to better reflect the complex processes underlying attention control. In addition, model comparisons and integration with other approaches such as the one described in Section 5.1.7 remain to be explored in detail. The following paragraphs outline some

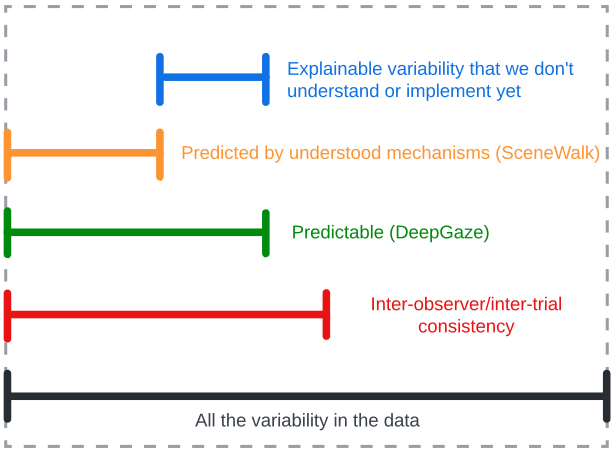

Figure 5.4 **Conceptual visualization** of the variance in eye movement data and the parts explained by the DeepGaze III and SceneWalk models. The difference between the performance of the models represents variance that is consistent over observers but not yet understood explicitly.

ideas that may represent future directions for the SceneWalk model.

Evidence suggests that the relative influence of bottom-up, top-down and systematic influences changes over the viewing duration (Schütt et al., 2019). While the first fixation is primarily guided by bottom-up factors, later parts of the scan path are predominantly influenced by higher-level factors. The SceneWalk model currently has a single set of parameters that characterize the behavior across the complete viewing duration. The only component of the SceneWalk model that is time dependent in this way, is the central fixation bias and its decay, which is added to only the first fixation in a sequence. In order to better reflect the complex processes underlying attention control, a future direction of the SceneWalk model could include fitting parameters that change over the viewing duration, perhaps using a hierarchical Bayesian approach.

As we showed in Section 5.1.7, the SceneWalk model predicts short saccades more consistently than long saccades. This indicates the absence of an effective Exploration-Exploitation mechanism. The use of a more heavy-tailed distribution for attention is a first step towards addressing this problem. This could be extended by assuming a set of distributions with a range of sizes and weights. An alternative approach takes inspiration from the Exploration-Exploitation paradigm (Malem-Shinitski et al., 2020) allowing the model to switch or vary between a broader and narrower focus over time.

The set of tasks that were modeled using SceneWalk, i.e., counting and guessing tasks performed on photographs of natural scenes, are only a few examples of the possible applications. SceneWalk could also applied to experiments that vary body posture (Backhaus & Engbert, 2022a). Another interesting future direction for SceneWalk could be to investigate the significant viewing differences between viewing familiar and unfamiliar images (Kaspar & König, 2011a; Kaspar & König, 2011b). We collected an extensive appropriate data set for this purpose, which has been partially

published (Schwetlick, Backhaus et al., 2020).

5.2 Insights from modeling fixational eye movement

While the purpose of macroscopic eye movements is well-understood, i.e., information is acquired during fixations and saccades shift the high acuity fovea to attended regions, the purpose of fixational eye movements is more contentious. The leading hypothesis is that the principal job of fixational eye movements is to counteract visual fading due to neural adaptation (Coppola & Purves, 1996; Martinez-Conde et al., 2004; Martinez-Conde et al., 2006). Moreover, fixational eye movement has also been found to contribute to high acuity vision (Intoy & Rucci, 2020) and is related to attentional processes as well as oculomotor control (Engbert & Kliegl, 2003; Hafed & Clark, 2002). In Chapter 4, we implemented a dynamical model of ocular drift as a likelihood-based model and explored the statistical properties of the trajectories as well as individual differences. The model also shows evidence of a relationship between ocular drift and microsaccades. In the following section I embed our findings into the literature and discuss the implications.

5.2.1 Individual differences

There is considerable individual variation in observed fixational eye movements, both in terms of ocular drift and microsaccades (Cherici et al., 2012; Poynter et al., 2013). This individual variability is often disregarded by taking population averages in an attempt to understand the common generating process. However, investigating the differences between individuals and the axes along which they occur, can also provide valuable insights. Differences in the expression of certain tendencies could be related to other features of the individual's visual system, as in the case of visual acuity (Cherici et al., 2012). Additionally, differences may imply the structure of the underlying mechanisms. Systematic variation in a specific property indicates that this property must be represented in the system.

In Chapter 4, we showed that the parameters of the SAW model, are able to partly capture and replicate these individual differences. To our knowledge, this study represents the first attempt to model individual variation in fixational eye movements at this level of detail. We find that the parameters of the model converge to distinctly different values for the different participants and that using these parameters in posterior predictive checks also allows the model to generate individually characteristic data. The success of the model in capturing individual variation lends support to its biological plausibility. We conclude that the implemented model mechanisms, i.e., self-avoidance mechanism, a stepping distribution and a confining potential, are a plausible basis for generating the observed fixational eye movement behavior and individual differences.

Individual differences in fixational eye movements have been found to be consistently characteristic to a point where they can be used in biometric identification. This

identification application is discussed for macroscopic eye movements in Section 5.1.4. Typically, viewer identification extracts and uses macroscopic eye movements, i.e., fixations and saccades. Jäger et al. (2020) find that a DNN model which takes the raw eye movement signal improves classification performance and speed. This can be interpreted as further evidence for the highly characteristic nature of fixational eye movements. It should be noted, however, that the amount of data and thereby extractable variance, is also much increased when considering millisecond-level data, as compared with fixation-level data. Potential future research could analyse the resultant DNN to investigate which of features it identifies as particularly important for the classification. Analysis of fitted DNNs is not trivial, but promising advances in this field can further our understanding of the relevant axes along which individuals differ.

In conclusion, our investigation of individual differences in fixational eye movements sheds light on the variability that exists between individuals in terms of ocular drift and microsaccades. The success of the model in capturing individual variation lends support to its biological plausibility and potential relevance for understanding the nature and function of fixational eye movements. Overall, our findings highlight the importance of considering individual differences in the study of fixational eye movements.

5.2.2 The relationship of drift and microsaccades

The relationship between different kinds of fixational eye movement remains a topic of debate. Chapter 4 provides evidence supporting the connection of fixational drift and microsaccade triggers: a build-up of activation at the current position of the model is associated with the occurrence of microsaccades. However, the exact mechanism behind this relationship and the potential causal direction between activation and microsaccades requires further investigation. In this section I discuss how the SAW model and its findings about microsaccades relate to three hypothesized functions of fixational eye movement: fading prevention, fixation control and visual acuity.

Prevention of fading has long been recognized as a primary role of fixational eye movement. Early studies considered fixational eye movements necessary due to the properties of retinal neurons, despite the additional complexity it introduces to the visual processing stream. More recent work revealed that later stages of visual processing also benefit from it (Rucci & Victor, 2015). In their role as a fading prevention mechanism, it has been suggested that microsaccades and drift fulfill subtly different but complementary roles. Drift continously prevents fading. By contrast, microsaccades restore vision once fading has occurred (McCamy et al., 2012). Consistent with this idea, Engbert and Mergenthaler (2006) found that microsaccades are associated with low retinal image slip, i.e., that drift motion is slower before microsaccade onset. The authors propose that this slowdown causally triggers microsaccades in order to prevent fading.

While early hypotheses suggested that microsaccades serve a corrective function for ocular drift (Cornsweet, 1956), studies have shown that microsaccades can be explo-

rative as well as corrective (Engbert & Kliegl, 2004). Both drift and microsaccades serve a function for precisely repositioning the eye for high acuity vision (Intoy & Rucci, 2020). In the SAW model the eye is constrained by a potential, with higher activation at greater distances to the intended fixation position. The idea that microsaccades are associated with higher activation in the model is consistent with the concept of corrective microsaccades. In fact, the model comparisons suggest that the activation increase before microsaccades is primarily related to the confining potential.

Recent research has demonstrated that microsaccades enhance the visibility of peripheral stimuli, facilitate high acuity vision, and are responsive to task demands (Intoy & Rucci, 2020; Ko et al., 2010; Rucci et al., 2007). Similarly drift movements were found to improve visual acuity, both in direct experimental studies (Intoy & Rucci, 2020) and by showing the benefit of introducing motion to models of feature detection (Schmittwilken & Maertens, 2022). These findings imply two interpretations. Both are consistent with the active vision framework and they are not mutually exclusive. First, motion benefits low level feature extraction. Second, an active cognitive control of fixational eye movements based on intention, task, and attention guide the movement in a top-down way. The SAW model represents the spatial characteristics of drift motion, but is stimulus-independent and does not implement any top-down control. Relating to the former statement, experimental work also shows that the individual drift motion is related to the visual acuity of the individual (Cherici et al., 2012). The individual variability captured by the SAW model may be related to individual differences in physiology and visual acuity. On the other hand, the role of active cognitive control in fixational eye movements may partly counteract the more systematic components. For example, the microsaccade rate has been found to decrease in high acuity vision tasks (Bridgeman & Palca, 1980) or else to be tuned to optimize task performance (Intoy & Rucci, 2020; Ko et al., 2010). As a result, situational influences of top-down control modulate the essential components of fixational eye movement.

5.2.3 Future directions in fixational eye movement research

The many recent publications studying fixational eye movements cited in this thesis reflect a growing interest in its functional role in visual processing. Advancements in eye-tracking technology are allowing researchers to study fixational eye movements in greater detail, leading to a better understanding of their nature and function. Additionally, instead of being viewed as the noisy result of the oculomotor system, or an inconvenient necessity for fading prevention, fixational eye movement is being recognized as a potentially significant contributor to vision. This shift in perspective has prompted increased attention particularly from researchers in the field of computer vision. A functional role in the human visual system may have significant implications for the development of computer vision technologies.

Recent research has shown the benefit of fixational eye movements particularly for high acuity vision (Intoy & Rucci, 2020; Rucci et al., 2007). It is likely that further research in this direction will find more specific correspondences between movement and perception. Presently, the scientific consensus has shifted from claiming that any

movement is a nuisance, to the idea that any movement can be an advantage. The logical next step is to ask whether the *specific* observed spatial statistics of ocular drift, is advantageous. A model presented by A. G. Anderson et al. (2020) uses Bayesian inference to simultaneously infer the movement and the stimulus. A generative version of this model may provide interesting insights into how movement may assist the perception. Another approach may be to use a model that successfully implements drift to the benefit of some sub-process of the visual system, like edge-detection (Schmittwilken & Maertens, 2022), and compare the effect of different sorts of movement (e.g., generated by the SAW model) on the model performance.

Experimental findings concerning the role of fixational eye movements for visual acuity as well as research into its relationship with attention (Engbert & Kliegl, 2003) emphasize the functional connection between perception and action. The movement depends on the stimulus, and the perception of the stimulus depends on the movement, in a highly integrative way. Therefore, a caveat of the SAW model is that it represents drift movement in the absence of specific input. A future extension of the model could include stimulus information, in a similar way as in the SceneWalk model. Another approach may be to fit the SAW model to different stimuli or different tasks, in order to analyze to which extent the stimulus dependence is systematic.

Lastly, the presented work establishes a strong base for researching the relationship between ocular drift and microsaccades, as discussed in Section 5.2.2. A combined model of drift and microsaccades based on a common activation representation would provide a cohesive framework for understanding what drives fixational eye movement and allow a more concrete investigation of the causal relationships. Overall, the application of the dynamical modeling framework offers many variations which will contribute to our understanding of visual processing.

6 General Discussion

At least I know I'm bewildered about the really fundamental and important facts of the universe.

Sir Terry Pratchett

6.1 Dynamical cognitive modeling

In this thesis I developed two models of human eye movement using a dynamical, process-oriented, and biologically plausible approach. Dynamical models have been shown to be effective in modeling a wide range of cognitive processes, from movement preparation to decision-making, and sensorimotor integration. A key advantage of this approach is that it allows for investigating the underlying mechanisms that produce and constrain observable behavior. Specifically, it highlights the relevance of microscopic physiological-, neural-, and kinematic properties for high-level behavior. This is particularly apparent in Chapter 2, which presents a substantial improvement including low-level attentional mechanisms in the model.

Additionally, cognitive processes typically contain a large amount of individual variation. The presented approach allows us to investigate vision at the level of individual subjects, by analyzing the differences in the fitted parameters of the model (Chapters 2, 3, and 4). Moreover, the same methods may be applied to modeling differences in conditions within subject, to better understand variation caused, for example, by task (Chapter 4).

The suitability of the dynamical data assimilation approach is exemplified by two models of different types of eye movement by explicitly representing the action space and using differential equations to define its evolution over time. The resulting model predictions can be compared to the corresponding empirical observations, providing an opportunity to test the specific assumptions against the data. In this section, I evaluate the methodology and highlight its particular suitability for modeling cognitive processes.

6.1.1 Dynamic modeling for dynamic processes

Conceptualizing eye movements as an active and dynamic process is crucial to understanding visual perception and its underlying mechanisms. The interaction of perception, i.e., the processing of signals that begin in photoreceptor cells in the retina, and action, i.e., the motion of the eyes which moves input across the retina, is a complex and dynamical process which cannot be separated. Movement is not just a quirk of the visual system, but a core component, which takes a variety of roles, from fading prevention to scene exploration. The precise visual input to the cells is constantly changing, while simultaneously cognition in the brain evolves, partly in response to the input and partly because other sensations and thoughts may occur in parallel. Thus, vision is an active process, which must be modeled and understood under these constraints.

The dynamical modeling framework and the associated data assimilation techniques applied in this thesis allow us to leverage the full information present in time-ordered data. Data assimilation involves using a mathematical model in order to estimate the current state of a system and its future behavior. This approach emphasizes the dynamical nature of the process and provides insights into the time-course of cognitive processes involved in visual processing, decision-making, and attention allocation. The advantages are two-fold: First, taking into account the dynamical nature of the system in the model architecture allows us investigate specific mechanisms that produce and constrain observable behavior. Second, we use each event in the sequence to provide better predictions of future events and to understand the implications of dependencies over time. Overall, this allows for a more detailed understanding of the cognitive mechanisms underlying eye movements and can lead to the development of more accurate models.

6.1.2 Likelihood and Bayesian parameter inference

Models in the field of Psychology and, to a lesser extent in cognitive science, have historically lacked a consistent and statistically rigorous framework (Bechtel & Abrahamsen, 2010; Cummins, 2010). When mathematical models are suggested, frequently the proof of concept is deemed sufficient, or else the parameter fitting process is undocumented (e.g., in Itti et al., 1998, or Boccignone and Ferraro, 2004). Parameter fitting procedures, where used and communicated, typically involve the use of a loss function (e.g., Engbert, Trukenbrod, et al., 2015; Jarodzka et al., 2010). The loss is defined as a metric that quantifies the fit between simulated- and experimentally observed data using a broad range of statistics (Le Meur & Baccino, 2012). The very serious disadvantage of this is that the choice of the statistics is arbitrary and different authors will likely apply different criteria, making an fair comparison of models difficult. It is likely that each model will perform well in the specific metrics it was fitted to and not others (Schütt et al., 2017).

Instead, as suggested by Schütt et al. (2017), in the work presented in this thesis we use the likelihood, i.e., the probability of the model given the data, as a global

and unbiased metric, in order to find parameters for the proposed models. It provides a global measure of how well the model fits the experimental data. As described in Section 1.2.1, likelihood-based modeling is at the center of the data-assimilation framework, opening the doors to rigorous parameter inference and model comparisons.

For both models presented in this thesis likelihood computation is straightforward: both models are deterministic and were implemented with likelihood computation in mind. For more complex models, approximate methods can be used to preserve the benefits of the likelihood approach (Seelig et al., 2020). In the context of static saliency prediction and model benchmarking, Kümmerer et al. (2015) apply a post-hoc likelihood approximation to models that are not explicitly likelihood-based. Here we use a Bayesian approach to compute full marginal posteriors for each parameter. As compared to an individual point estimate, a posterior distribution contains valuable information about the parameter variability and how well the data constrains the model with regard to specific components.

In a dynamical model, the likelihood is computed for each event in the time-ordered sequence, where each model state is based on the preceding events. In order to infer parameters, the event-likelihoods are summed to obtain the likelihood of the whole dataset. Additionally, this event-level information can itself be used as an analysis tool: in Section 5.1.7 we qualitatively analyse specific situations where model performance is particularly good or bad to draw inspiration for missing mechanisms.

It is important to keep in mind that the model likelihood is a very general measure of performance, which does not address specific statistical properties. Therefore, it is necessary to use other statistical methods and metrics in addition to likelihood to identify which aspects of the data are well-modeled and which are not. Overall, this thesis demonstrates the value of a mechanistic, biologically plausible approach to understanding the underlying processes that give rise to human eye movements.

6.1.3 Individual differences

A key feature of the results presented here using a likelihood-based cognitive modeling framework is its ability to capture individual differences through variation in model parameters. Individual variation has a dominant influence on eye movement data and on experimental data in cognitive science in general (Bargary et al., 2017; Kliegl, 2010). However, for models of general cognitive function individual differences are rarely considered. One reason for this may be that non-optimal fitting procedures have high data requirements and are very computationally expensive. The presented work addresses both challenges and successfully captures individual differences, and even task differences within individuals.

The amount of data needed to fit a model varies greatly with the number of parameters, with the quality of the model, and the fitting method. Compared to many data-driven models, such as DNNs, dynamical models tend to have a limited number of parameters. Combined with an optimized and highly parallelized computational implementation, as well as an equally optimized fitting algorithm, we found the fitting of models at the individual level to be feasible. However, it is important to note

that the work in this thesis is based on many thousands, if not millions, of hours of computing time on powerful compute-clusters. Thus, the computational challenges of modeling individual variation can be addressed but should not be underestimated.

Modeling individual differences can provide important insights into the nature of the underlying data-generating processes. As we model behavior in a hypothesis-driven way, the behavioral differences between subjects translate to interpretable differences in the space of parameters. Using the model to generate data, shows that individual differences in behavior can be captured by the model in both the SceneWalk and SAW models. The variation captured by the model is capable of causally explaining individual differences in behavior. This reinforces the biological plausibility of the model, as it shows that the variation between individuals corresponds to parameterized mechanisms in the model. By capturing the idiosyncrasies of individual behavior, we can gain insights into the structure and function of the visual system at a level of detail that is not possible when considering only population averages.

Here, we fitted individual models by simply separating the data sets and running the appropriate fitting procedures one-by-one. Another method for integrating over differences between participants are hierarchical Bayesian models. A hierarchical model typically implements the individual differences in parameters following an additional model for the distribution of parameters. A first advance into using hierarchical models in the context of the SceneWalk model was made by Schütt et al. (2017). A model comparison of an averaged- versus an individually fitted SAW model indicates that a hierarchical approach may be beneficial. Hierarchical Modeling stabilizes the parameter estimates, particularly for subjects where the model is not well-constrained by the data. Implementing such an approach for the current version of SceneWalk or for the SAW model, could be an interesting next step for investigating individual differences, particularly where sufficiently large data sets are not available.

6.2 Final conclusion

Due to its complexity and high relevance for interacting with the world, the field of vision science is relevant not only for understanding perception and action, but also for applications of machine vision. In this thesis I applied dynamical modeling and the data assimilation framework to the field of eye movement research. As the nature of visual perception is fundamentally tied to eye movement and the underlying decision-making processes, dynamical modeling is a particularly suitable approach. The way perception and action change and interact over time is represented in dynamical models as a set rules for how model states evolve over time. As a result, the current state is computed by taking into account all past states. In the presented examples, these rules are implemented in a biologically plausible way, yielding a process-oriented, hypothesis-driven model which allows a detailed exploration of the processes underlying eye movement.

In conclusion, cognitive dynamical modeling and the related statistical and mathematical methods presented here are highly seminal. Although the scope of this thesis is

limited, it may serve as an example of how data assimilation techniques can be applied to models of vision and to cognitive models in general. Certainly, further development of these methods will greatly benefit our understanding of how basic mechanisms of perception and action interact to produce rich, complex and creative behavior.

Bibliography

Adeli, H., Vitu, F., & Zelinsky, G. J. (2016). A model of the superior colliculus predicts fixation locations during scene viewing and visual search. *Journal of Neuroscience*, *37*(6), 1453–1467. https://doi.org/10.1523/jneurosci.0825-16.2016

Adler, F. H., & Fliegelman, M. (1934). Influence of fixation on the visual acuity. *Archives of Ophthalmology*, *12*(4), 475–483. https://doi.org/10.1001/archopht.1934.00830170013002

Alexander, R. G., & Martinez-Conde, S. (2019). Eye movement research. In C. Klein & U. Ettinger (Eds.), *Eye movement research* (pp. 73–115). Springer International Publishing. https://doi.org/10.1007/978-3-030-20085-5_3

Aloimonos, Y., & Rosenfeld, A. (1991). Computer vision. *Science*, *253*(5025), 1249–1254. https://doi.org/10.1126/science.1891713

Anderson, A. G., Ratnam, K., Roorda, A., & Olshausen, B. A. (2020). High-acuity vision from retinal image motion. *Journal of Vision*, *20(7):34*. https://doi.org/10.1167/jov.20.7.34

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Winston and Sons.

Asch, M., Bocquet, M., & Nodet, M. (2016). *Data assimilation: Methods, algorithms, and applications*. Society for Industrial and Applied Mathematics.

Backhaus, D., & Engbert, R. (2022a). Investigating the effects of task and body movement on the generalizability of scene viewing experiments. In V. McGowan, A. Pagán, K. B. Paterson, D. Souto, & R. Groner (Eds.), *Book of abstracts of the 21st european conference on eye movements*. Journal of Eye Movement Research, 15 *(5)*. https://doi.org/10.16910/jemr.15.5.1

Backhaus, D., & Engbert, R. (2022b). *Scene viewing in laboratory experiments: How "free viewing" task and a chin rest influence eye movements* [Data Set]. OSF. https://osf.io/yaqgz

Backhaus, D., Engbert, R., Rothkegel, L. O. M., & Trukenbrod, H. A. (2020). Task-dependence in scene perception: Head unrestrained viewing using mobile eye-tracking. *Journal of Vision*, *20(5):3*. https://doi.org/10.1167/jov.20.5.3

Bahill, A. T., Clark, M. R., & Stark, L. (1975). The main sequence, a tool for studying human eye movements. *Mathematical Biosciences*, *24*(3), 191–204. https://doi.org/10.1016/0025-5564(75)90075-9

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, *20*(4), 723–742. https://doi.org/10.1017/S0140525X97001611

Bargary, G., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Hogg, R. E., & Mollon, J. D. (2017). Individual differences in human eye movements: An oculomotor signature? *Vision Research*, *141*, 157–169. https://doi.org/10.1016/j.visres.2017.03.001

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious mixed models* [Preprint]. ArXiv. https://doi.org/10.48550/arXiv.1506.04967

Bays, P. M., & Husain, M. (2012). Active inhibition and memory promote exploration and search of natural scenes. *Journal of Vision*, *12(8):8*. https://doi.org/10.1167/12.8.8

Bear, M. F., Connors, B. W., & Paradiso, M. A. (2007). *Neuroscience* (S. Katz, Ed.). Lippincott Williams & Wilkins.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *36*(2), 421–441. https://doi.org/10.1016/j.shpsc.2005.03.010

Bechtel, W., & Abrahamsen, A. (2010). Dynamic mechanistic explanation: Computational modeling of circadian rhythms as an exemplar for cognitive science. *Studies in History and Philosophy of Science Part A*, *41*(3), 321–333. https://doi.org/10.1016/j.shpsa.2010.07.003

Becker, W., & Jürgens, R. (1979). An analysis of the saccadic system by means of double step stimuli. *Vision Research*, *19*(9), 967–983. https://doi.org/10.1016/0042-6989(79)90222-0

Ben-Shushan, N., Shaham, N., Joshua, M., & Burak, Y. (2022). Fixational drift is driven by diffusive dynamics in central neural circuitry. *Nature Communications*, *13*, 1697. https://doi.org/10.1038/s41467-022-29201-y

Bindemann, M. (2010). Scene and screen center bias early eye movements in scene viewing. *Vision Research*, *50*(23), 2577–2587. https://doi.org/10.1016/j.visres.2010.08.016

Bisley, J. W., & Mirpour, K. (2019). The neural instantiation of a priority map. *Current Opinion in Psychology*, *29*, 108–112. https://doi.org/10.1016/j.copsyc.2019.01.002

Blakemore, C. T., & Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology*, *203*(1), 237–260. https://doi.org/10.1113/jphysiol.1969.sp008862

Boccignone, G., & Ferraro, M. (2004). Modelling gaze shift as a constrained random walk. *Physica A: Statistical Mechanics and Its Applications*, *331*(1-2), 207–218. https://doi.org/10.1016/j.physa.2003.09.011

Boi, M., Poletti, M., Victor, J. D., & Rucci, M. (2017). Consequences of the oculomotorcycle for the dynamics of perception. *Current Biology*, *27*(9), 1268–1277. https://doi.org/10.1016/j.cub.2017.03.034

Boisvert, J. F. G., & Bruce, N. D. B. (2016). Predicting task from eye movements: On the importance of spatial distribution, dynamics, and image features. *Neurocomputing*, *207*, 653–668. https://doi.org/10.1016/j.neucom.2016.05.047

Bonev, B., Chuang, L. L., & Escolano, F. (2013). How do image complexity, task demands and looking biases influence human gaze behavior? *Pattern Recognition Letters*, *34*(7), 723–730. https://doi.org/10.1016/j.patrec.2012.05.007

Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(1), 185–207. https://doi.org/10.1109/tpami.2012.89

Borji, A., & Itti, L. (2014). Defending yarbus: Eye movements reveal observers' task. *Journal of Vision*, *14(3):29*. https://doi.org/10.1167/14.3.29

Borji, A., & Itti, L. (2015). *Cat2000: A large scale fixation dataset for boosting saliency research* [Preprint]. ArXiv. https://doi.org/10.48550/arXiv.1505.03581

Borji, A., Sihite, D. N., & Itti, L. (2013). Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. *IEEE Transactions on Image Processing*, *22*(1), 55–69. https://doi.org/10.1109/tip.2012.2210727

Bowers, N. R., Gautier, J., Lin, S., & Roorda, A. (2021). *Fixational eye movements depend on task and target* [Preprint]. BioArXiv. https://doi.org/10.1101/2021.04.14.439841

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations (and discussion). *Journal of the Royal Statistical Society: Series B (methodological)*, *26*(2), 211–252. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Braak, C. J. F. T. (2006). A markov chain monte carlo version of the genetic algorithm differential evolution: Easy bayesian computing for real parameter spaces. *Statistics and Computing*, *16*(3), 239–249. https://doi.org/10.1007/s11222-006-8769-1

Bridgeman, B., & Palca, J. (1980). The role of microsaccades in high acuity observational tasks. *Vision Research*, *20*(9), 813–817. https://doi.org/10.1016/0042-6989(80)90013-9

Brodersen, K. H., Penny, W. D., Harrison, L. M., Daunizeau, J., Ruff, C. C., Duzel, E., Friston, K. J., & Stephan, K. E. (2008). Integrated bayesian models of learning and decision making for saccadic eye movements. *Neural Networks*, *21*(9), 1247–1260. https://doi.org/10.1016/j.neunet.2008.08.007

Brooks, S., Gelman, A., Jones, G., & Meng, X.-L. (Eds.). (2011). *Handbook of markov chain monte carlo*. Chapman; Hall/CRC. https://doi.org/10.1201/b10905

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). *Language models are few-shot learners* [Preprint]. ArXiv. https://doi.org/10.48550/arXiv.2005.14165

Bruce, N. D. B., & Tsotsos, J. K. (2009). Saliency, attention, and visual search: An information theoretic approach. *Journal of Vision*, *9(3):5*. https://doi.org/10.1167/9.3.5

Burak, Y., Rokni, U., Meister, M., & Sompolinsky, H. (2010). Bayesian model of dynamic image stabilization in the visual system. *Proceedings of the National Academy of Sciences*, *107*(45), 19525–19530. https://doi.org/10.1073/pnas.1006076107

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*(3), 432–459. https://doi.org/10.1037/0033-295x.100.3.432

Buswell, G. T. (1935). *How people look at pictures: A study of the psychology and perception in art.* University of Chicago Press.

Bylinskii, Z., Judd, T., Borji, A., Itti, L., Durand, F., Oliva, A., & Torralba, A. (2015). *MIT saliency benchmark* [Website]. http://saliency.mit.edu/

Bylinskii, Z., Recasens, A., Borji, A., Oliva, A., Torralba, A., & Durand, F. (2016). Where should saliency models look next? In *Computer vision – ECCV 2016* (pp. 809–824). Springer International Publishing. https://doi.org/10.1007/978-3-319-46454-1_49

Cadena, S. A., Denfield, G. H., Walker, E. Y., Gatys, L. A., Tolias, A. S., Bethge, M., & Ecker, A. S. (2019). Deep convolutional models improve predictions of macaque v1 responses to natural images (W. Einhäuser, Ed.). *PLoS Computational Biology*, *15*(4), e1006897. https://doi.org/10.1371/journal.pcbi.1006897

Carpenter, R. H. S. (2000). The neural control of looking. *Current Biology*, *10*(8), R291–R293. https://doi.org/10.1016/s0960-9822(00)00430-9

Carpenter, R. H. S., & Reddi, B. A. J. (2001). Reply to 'putting noise into neurophysiological models of simple decision making'. *Nature Neuroscience*, *4*(4), 337–337. https://doi.org/10.1038/85960

Cartwright, N. (1999). *The dappled world: A study of the boundaries of science*. Cambridge University Press.

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, *9(3):6*. https://doi.org/10.1167/9.3.6

Castet, E., & Masson, G. S. (2000). Motion perception during saccadic eye movements. *Nature Neuroscience*, *3*(2), 177–183. https://doi.org/10.1038/72124

Castet, E., & Montagnini, A. (2007). Spatiotemporal dynamics of visual attention during saccade preparation: Independence and coupling between attention and movement planning. *Journal of Vision*, *7(14):8*. https://doi.org/10.1167/7.14.8

Cerf, M., Harel, J., Einhäuser, W., & Koch, C. (2008). Predicting human gaze using low-level saliency combined with face detection. In D. Koller (Ed.), *Advances in neural information processing systems* (pp. 241–248). MIT Press.

Chapman, P., Underwood, G., & Roberts, K. (2002). Visual search patterns in trained and untrained novice drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, *5*(2), 157–167. https://doi.org/10.1016/s1369-8478(02)00014-1

Chen, Y., Yang, Z., Ahn, S., Samaras, D., Hoai, M., & Zelinsky, G. (2020). *COCO-search18: A dataset for predicting goal-directed attention control* [Preprint]. BioArXiv. https://doi.org/10.1101/2020.07.27.221499

Cherici, C., Kuang, X., Poletti, M., & Rucci, M. (2012). Precision of sustained fixation in trained and untrainedobservers. *Journal of Vision*, *12(6):31*. https://doi.org/10.1167/12.6.31

Churchland, P. M. (1989). Some reductive strategies in cognitive neurobiology. In *Rerepresentation* (pp. 223–253). Springer Netherlands. https://doi.org/10.1007/978-94-009-2649-3_12

Clarke, A. D. F., Hunt, A. R., & Hughes, A. (2021). *Building bayesian cognitive models of visual foraging* [Preprint]. OSF. https://doi.org/10.31234/osf.io/zacr9

Clarke, A. D. F., Stainer, M. J., Tatler, B. W., & Hunt, A. R. (2017). The saccadic flow baseline: Accounting for image-independent biases in fixation behavior. *Journal of Vision*, *17(11):12*. https://doi.org/10.1167/17.11.12

Collaboration, O. S. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Collewijn, H., Erkelens, C. J., & Steinman, R. M. (1988). Binocular co-ordination of human horizontal saccadic eye movements. *Journal of Physiology*, *404*(1), 157–182. https://doi.org/10.1113/jphysiol.1988.sp017284

Collewijn, H., & Kowler, E. (2008). The significance of microsaccades for vision and oculomotor control. *Journal of Vision*, *8(14):20*. https://doi.org/10.1167/8.14.20

Collins, J. J., & De Luca, C. J. (1995). The effects of visual input on open-loop and closed-loop postural control mechanisms. *Experimental Brain Research*, *103*(1), 151–163. https://doi.org/10.1007/BF00241972

Coppola, D., & Purves, D. (1996). The extraordinarily rapid disappearance of entopic images. *Proceedings of the National Academy of Sciences*, *93*(15), 8001–8004. https://doi.org/10.1073/pnas.93.15.8001

Cornelissen, T. H. W., & Võ, M. L.-H. (2017). Stuck on semantics: Processing of irrelevant object-scene inconsistencies modulates ongoing gaze behavior. *Attention, Perception, & Psychophysics*, *79*(1), 154–168. https://doi.org/10.3758/s13414-016-1203-7

Cornsweet, T. N. (1956). Determination of the stimuli for involuntary drifts and saccadic eye movements10.1016/0042-6989(76)90156-5. *Journal of the Optical Society of America*, *46*(11), 987. https://doi.org/10.1364/josa.46.000987

Costela, F. M., Otero-Millan, J., McCamy, M. B., Macknik, S. L., Troncoso, X. G., Jazi, A. N., Crook, S. M., & Martinez-Conde, S. (2014). Fixational eye movement correction of blink-induced gaze position errors (K. Paterson, Ed.). *PLoS ONE*, *9*(10), e110889. https://doi.org/10.1371/journal.pone.0110889

Coutrot, A., Hsiao, J. H., & Chan, A. B. (2017). Scanpath modeling and classification with hidden markov models. *Behavior Research Methods*, *50*, 362–379. https://doi.org/10.3758/s13428-017-0876-8

Crüwell, S., & Evans, N. J. (2019). *Preregistration in complex contexts: A preregistration template for the application of cognitive models* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/2hykx

Cummins, R. (2010). 'How does it work?' vs. 'What are the laws?' In F. Keil & R. Wilson (Eds.), *The world in the head* (pp. 282–310). Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199548033.003.0016

De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, *22*(5), 545–559. https://doi.org/10.1016/0042-6989(82)90113-4

DeAngelus, M., & Pelz, J. B. (2009). Top-down control of eye movements: Yarbus revisited. *Visual Cognition*, *17*(6-7), 790–811. https://doi.org/10.1080/13506280902793843

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*(12), 1827–1837. https://doi.org/10.1016/0042-6989(95)00294-4

Ditchburn, R. W., Fender, D. H., & Mayne, S. (1959). Vision with controlled movements of the retinal image. *Journal of Physiology*, *145*(1), 98–107. https://doi.org/10.1113/jphysiol.1959.sp006130

Ditchburn, R. W., & Ginsborg, B. L. (1952). Vision with a stabilized retinal image. *Nature*, *170*(4314), 36–37. https://doi.org/10.1038/170036a0

Donner, K., & Hemilä, S. (2007). Modelling the effect of microsaccades on retinal responses to stationary contrast patterns. *Vision Research*, *47*(9), 1166–1177. https://doi.org/10.1016/j.visres.2006.11.024

Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision*, *8(14):18*. https://doi.org/10.1167/8.14.18

Einhäuser, W., & Nuthmann, A. (2016). Salient in space, salient in time: Fixation probability predicts fixation duration during natural scene viewing. *Journal of Vision*, *16(11):13*. https://doi.org/10.1167/16.11.13

Einhäuser, W., Rutishauser, U., Koch, C., et al. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision*, *8(2):2*. https://doi.org/10.1167/8.2.2

Eizenman, M., Hallett, P. E., & Frecker, R. C. (1985). Power spectra for ocular drift and tremor. *Vision Research*, *25*(11), 1635–1640. https://doi.org/10.1016/0042-6989(85)90134-8

Engbert, R. (2012). Computational modeling of collicular integration of perceptual responses and attention in microsaccades. *Journal of Neuroscience*, *32*(23), 8035–8039. https://doi.org/10.1523/jneurosci.0808-12.2012

Engbert, R., & Mergenthaler, K. (2006). Microsaccades are triggered by low retinal image slip. *Proceedings of the National Academy of Sciences*, *103*(18), 7192–7197. https://doi.org/10.1073/pnas.0509557103

Engbert, R., Sinn, P., Mergenthaler, K., & Trukenbrod, H. A. (2015). *Microsaccade toolbox* (Version R) [Software]. http://read.psych.uni-potsdam.de/attachments/article/140/MS_Toolbox_R.zip

Engbert, R., Trukenbrod, H. A., Barthelme, S., & Wichmann, F. A. (2015). Spatial statistics and attentional dynamics in scene viewing. *Journal of Vision*, *15(1):14*. https://doi.org/10.1167/15.1.14

Engbert, R. (2006). Microsaccades: A microcosm for research on oculomotor control, attention, and visual perception. *Progress in Brain Research*, *154*, 177–192. https://doi.org/10.1016/s0079-6123(06)54009-9

Engbert, R. (2021). *Dynamical Models in Neurocognitive Psychology*. Springer Nature Publishing. https://doi.org/10.1007/978-3-030-67299-7

Engbert, R., & Kliegl, R. (2001). Mathematical models of eye movements in reading: A possible role for autonomous saccades. *Biological Cybernetics*, *85*(2), 77–87. https://doi.org/10.1007/pl00008001

Engbert, R., & Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision Research*, *43*(9), 1035–1045. https://doi.org/10.1016/s0042-6989(03)00084-1

Engbert, R., & Kliegl, R. (2004). Microsaccades keep the eyes' balance during fixation. *Psychological Science*, *15*(6), 431–431. https://doi.org/10.1111/j.0956-7976.2004.00697.x

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research*, *42*(5), 621–636. https://doi.org/10.1016/S0042-6989(01)00301-7

Engbert, R., Mergenthaler, K., Sinn, P., & Pikovsky, A. (2011). An integrated model of fixational eye movements and microsaccades. *Proceedings of the National Academy of Sciences*, *108*(39), 16149–16150. https://www.pnas.org/content/108/39/E765/1

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, *112*(4), 777–813. https://doi.org/10.1037/0033-295X.112.4.777

Engbert, R., Rabe, M. M., Schwetlick, L., Seelig, S. A., Reich, S., & Vasishth, S. (2022). Data assimilation in dynamical cognitive science. *Trends in Cognitive Sciences*, *26*(2), 99–102. https://doi.org/10.1016/j.tics.2021.11.006

Erlhagen, W., & Schöner, G. (2002). Dynamic field theory of movement preparation. *Psychological Review*, *109*(3), 545–572. https://doi.org/10.1037/0033-295x.109.3.545

Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198524793.001.0001

Findlay, J. M., & Walker, R. (1999a). How are saccades generated? *Behavioral and Brain Sciences*, *22*(4), 706–713. https://doi.org/10.1017/s0140525x99552151

Findlay, J. M., & Walker, R. (1999b). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences*, *22*(4), 661–674. https://doi.org/10.1017/S0140525X99002150

Foulsham, T., & Kingstone, A. (2010). Asymmetries in the direction of saccades during perception of scenes and fractals: Effects of image type and image features. *Vision Research*, *50*(8), 779–795. https://doi.org/10.1016/j.visres.2010.01.019

Foulsham, T., Kingstone, A., & Underwood, G. (2008). Turning the world around: Patterns in saccade direction vary with picture orientation. *Vision Research*, *48*(17), 1777–1790. https://doi.org/10.1016/j.visres.2008.05.018

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, *8(2):6*. https://doi.org/10.1167/8.2.6

Freund, H., & Grassberger, P. (1992). The red queen's walk. *Physica A: Statistical Mechanics and Its Applications*, *190*(3-4), 218–237. https://doi.org/10.1016/0378-4371(92)90033-m

Funke, C. M., Borowski, J., Stosio, K., Brendel, W., Wallis, T. S. A., & Bethge, M. (2020). *The notorious difficulty of comparing human and machine perception* [Preprint]. ArXiv. https://doi.org/10.48550/arXiv.2004.09406

Geirhos, R., Janssen, D. H. J., Schütt, H. H., Rauber, J., Bethge, M., & Wichmann, F. A. (2017). *Comparing deep neural networks against humans: Object recognition when the signal gets weaker* [Preprint]. ArXiv. https://doi.org/10.48550/arXiv.1706.06969

Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., & Brendel, W. (2018). *Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness*. ArXiv. https://doi.org/10.48550/arXiv.1811.12231

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. Taylor & Francis Ltd. https://doi.org/https://doi.org/10.1201/b16018

Gershenfeld, N. (1999). *The nature of mathematical modeling*. Cambridge University Press, Cambridge, U.K.

Gilchrist, I. D., & Harvey, M. (2006). Evidence for a systematic component within scan paths in visual search. *Visual Cognition*, *14*(4-8), 704–715. https://doi.org/10.1080/13506280500193719

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (Eds.). (1996). *Markov chain Monte Carlo in practice*. Chapman & Hall/CRC.

Goldbeter, A. (1995). A model for circadian oscillations in the drosophila period protein (PER). *Proceedings of the Royal Society, Series B: Biological Sciences*, *261*(1362), 319–324. https://doi.org/10.1098/rspb.1995.0153

Golomb, J. D., Chun, M. M., & Mazer, J. A. (2008). The native coordinate system of spatial attention is retinotopic. *Journal of Neuroscience*, *28*(42), 10654–10662. https://doi.org/10.1523/jneurosci.2525-08.2008

Golomb, J. D., Marino, A. C., Chun, M. M., & Mazer, J. A. (2010). Attention doesn't slide: Spatiotopic updating after eye movements instantiates a new, discrete attentional locus. *Attention, Perception, & Psychophysics*, *73*(1), 7–14. https://doi.org/10.3758/s13414-010-0016-3

Greene, M. R., Liu, T., & Wolfe, J. M. (2012). Reconsidering Yarbus: A failure to predict observers' task from eye movement patterns. *Vision Research*, *62*, 1–8. https://doi.org/10.1016/j.visres.2012.03.019

Hafed, Z. M., & Clark, J. J. (2002). Microsaccades as an overt measure of covert attention shifts. *Vision Research*, *42*(22), 2533–2545. https://doi.org/10.1016/S0042-6989(02)00263-8

Haji-Abolhassani, A., & Clark, J. J. (2014). An inverse Yarbus process: Predicting observers' task from eye movement patterns. *Vision Research*, *103*, 127–142. https://doi.org/10.1016/j.visres.2014.08.014

Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, *51*(5), 347–356. https://doi.org/10.1007/bf00336922

Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research*, *18*(10), 1279–1296. https://doi.org/10.1016/0042-6989(78)90218-3

Hanning, N. M., Szinte, M., & Deubel, H. (2019). Visual attention is not limited to the oculomotor range. *Proceedings of the National Academy of Sciences*, *116*(19), 9665–9670. https://doi.org/10.1073/pnas.1813465116

Harel, J., Koch, C., & Perona, P. (2006). Graph-based visual saliency. *Proceedings of the 19th International Conference on Neural Information Processing Systems*, 545–552.

Harris, C. M., & Wolpert, D. M. (2006). The main sequence of saccades optimizes speed-accuracy trade-off. *Biological Cybernetics*, *95*(1), 21–29. https://doi.org/10.1007/s00422-006-0064-x

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, *57*(1), 97–109. https://doi.org/10.1093/biomet/57.1.97

Haxby, J. V., Hoffman, E. A., & Gobbini, M. (2002). Human neural systems for face recognition and social communication. *Biological Psychiatry*, *51*(1), 59–67. https://doi.org/10.1016/s0006-3223(01)01330-0

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, *3(1):6*. https://doi.org/10.1167/3.1.6

Helo, A., Pannasch, S., Sirri, L., & Rämä, P. (2014). The maturation of eye movement behavior: Scene viewing characteristics in children and adults. *Vision Research*, *103*, 83–91. https://doi.org/10.1016/j.visres.2014.08.006

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

Henderson, J. M. (2011). Eye movements and scene perception. In S. Liversedge, I. Gilchrist, & S. Everling (Eds.), *The oxford handbook of eye movements* (pp. 593–606). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199539789.013.0033

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537–562). Elsevier. https://doi.org/10.1016/B978-008044980-7/50027-6

Henderson, J. M., & Hayes, T. R. (2017). Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nature Human Behaviour*, *1*(10), 743–747. https://doi.org/10.1038/s41562-017-0208-0

Henderson, J. M., & Hayes, T. R. (2018). Meaning guides attention in real-world scene images: Evidence from eye movements and meaning maps. *Journal of Vision*, *18(6):10*. https://doi.org/10.1167/18.6.10

Henderson, J. M., Hayes, T. R., Peacock, C. E., & Rehrig, G. (2019). Meaning and attentional guidance in scenes: A review of the meaning map approach. *Vision*, *3*(2), 19. https://doi.org/10.3390/vision3020019

Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 269–293). Amsterdam: Elsevier.

Henderson, J. M., & Hollingworth, A. (2003). Eye movements, visual memory, and scene representation. In M. A. Anderson & G. Rhodes (Eds.), *Perception of faces, objects, and scenes: Analytic and holistic processes* (pp. 356–383). Oxford University Press.

Henderson, J. M., Phillip A., J. W., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 210–228. https://doi.org/10.1037/0096-1523.25.1.210

Henderson, J. M., & Pierce, G. L. (2008). Eye movements during scene viewing: Evidence for mixed control of fixation durations. *Psychonomic Bulletin & Review*, *15*(3), 566–573. https://doi.org/10.3758/pbr.15.3.566

Henderson, J. M., & Smith, T. J. (2009). How are eye fixation durations controlled during scene viewing? Further evidence from a scene onset delay paradigm. *Visual Cognition*, *17*(6-7), 1055–1082. https://doi.org/10.1080/13506280802685552

Herrmann, C. J. J., Metzler, R., & Engbert, R. (2017). A self-avoiding walk with neural delays as a model of fixational eye movements. *Scientific Reports*, *7*, 12958. https://doi.org/10.1038/s41598-017-13489-8

Higgins, G. C., & Stultz, K. F. (1953). Frequency and amplitude of ocular tremor. *Journal of the Optical Society of America*, *43*(12), 1136. https://doi.org/10.1364/josa.43.001136

Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics*, *57*(6), 787–795. https://doi.org/10.3758/bf03206794

Hopfinger, J. B., & Mangun, G. R. (1998). Reflexive attention modulates processing of visual stimuli in human extrastriate cortex. *Psychological Science*, *9*(6), 441–447. https://doi.org/10.1111/1467-9280.00083

Illian, J., Illian, A., Stoyan, H., & Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*. John Wiley & Sons. https://doi.org/10.1002/9780470725160

Intoy, J., & Rucci, M. (2020). Finely tuned eye movements enhance visual acuity. *Nature Communications*, *11*, 795. https://doi.org/10.1038/s41467-020-14616-2

Irwin, D. E., & Gordon, R. D. (1998). Eye movements, attention and trans-saccadic memory. *Visual Cognition*, *5*(1-2), 127–155. https://doi.org/10.1080/713756783

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254–1259. https://doi.org/10.1109/34.730558

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10-12), 1489–1506. https://doi.org/10.1016/s0042-6989(99)00163-7

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*(3), 194–203. https://doi.org/10.1038/35058500

Jäger, L. A., Makowski, S., Prasse, P., Liehr, S., Seidler, M., & Scheffer, T. (2020). Deep eye-dentification: Biometric identification using micro-movements of the eye. In *Machine learning and knowledge discovery in databases* (pp. 299–314). Springer International Publishing. https://doi.org/10.1007/978-3-030-46147-8_18

Jarodzka, H., Holmqvist, K., & Nyström, M. (2010). A vector-based, multidimensional scanpath similarity measure. *Proceedings of the 2010 Symposium on Eye-tracking Research & Applications - ETRA '10*. https://doi.org/10.1145/1743666.1743718

Judd, T., Durand, F., & Torralba, A. (2012). *A benchmark of computational models of saliency to predict human fixations* [Technical Report]. http://hdl.handle.net/1721.1/68590

Judd, T., Ehinger, K., Durand, F., & Torralba, A. (2009). Learning to predict where humans look. *2009 IEEE 12th International Conference on Computer Vision*. https://doi.org/10.1109/iccv.2009.5459462

Kanan, C., Ray, N. A., Bseiso, D. N. F., Hsiao, J. H., & Cottrell, G. W. (2014). Predicting an observer's task using multi-fixation pattern analysis. *ETRA '14: Proceedings of the Symposium on Eye Tracking Research and Applications*, 287–290. https://doi.org/10.1145/2578153.2578208

Kaplan, D. T. (2009). *Statistical modeling: A fresh approach*. MacAlester College.

Kaspar, K., & König, P. (2011a). Viewing behavior and the impact of low-level image properties across repeated presentations of complex scenes. *Journal of Vision*, *11(13):26*. https://doi.org/10.1167/11.13.26

Kaspar, K., & König, P. (2011b). Overt attention and context factors: The impact of repeated presentations, image type, and individual motivation (J. Z. Tsien, Ed.). *PLoS ONE*, *6*(7), e21719. https://doi.org/10.1371/journal.pone.0021719

Kassner, M., Patera, W., & Bulling, A. (2014). Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. https://doi.org/10.1145/2638728.2641695

Kelling, S., Hochachka, W. M., Fink, D., Riedewald, M., Caruana, R., Ballard, G., & Hooker, G. (2009). Data-intensive science: A new paradigm for biodiversity studies. *BioScience*, *59*(7), 613–620. https://doi.org/10.1525/bio.2009.59.7.12

Kelly, D. (1962). Information capacity of a single retinal channel. *IRE Transactions on Information Theory*, *8*(3), 221–226. https://doi.org/10.1109/tit.1962.1057716

Kerkouri, M. A., Tliba, M., Chetouani, A., & Harba, R. (2021). Salypath: A deep-based architecture for visual attention prediction. *2021 IEEE International Conference on Image Processing (ICIP)*. https://doi.org/10.1109/icip42928.2021.9506295

Kienzle, W., Franz, M. O., Schölkopf, B., & Wichmann, F. A. (2009). Center-surround patterns emerge as optimal predictors for human saccade targets. *Journal of Vision*, *9(5):7*. https://doi.org/10.1167/9.5.7

Killian, N. J., Jutras, M. J., & Buffalo, E. A. (2012). A map of visual space in the primate entorhinal cortex. *Nature*, *491*(7426), 761–764. https://doi.org/10.1038/nature11587

Klein, R. M. (2000). Inhibition of return. *Trends in Cognitive Sciences*, *4*(4), 138–147. https://doi.org/10.1016/S1364-6613(00)01452-2

Klein, R. M., & MacInnes, W. J. (1999). Inhibition of return is a foraging facilitator in visual search. *Psychological Science*, *10*(4), 346–352. https://doi.org/10.1111/1467-9280.00166

Kliegl, R. (2010). Experimental effects and individual differences in linear mixed models: Estimating the relationship between spatial, object, and attraction effects in visual attention. *Frontiers in Psychology*, *1*. https://doi.org/10.3389/fpsyg.2010.00238

Ko, H.-k., Poletti, M., & Rucci, M. (2010). Microsaccades precisely relocate gaze in a high visual acuity task. *Nature Neuroscience*, *13*(12), 1549–1553. https://doi.org/10.1038/nn.2663

Ko, H.-k., Snodderly, D. M., & Poletti, M. (2016). Eye movements between saccades: Measuring ocular drift and tremor. *Vision Research*, *122*, 93–104. https://doi.org/10.1016/j.visres.2016.03.006

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. In *Matters of intelligence* (pp. 115–141). Springer. https://doi.org/10.1007/978-94-009-3833-5_5

Kolb, B., Whishaw, I. Q., & Teskey, G. C. (2001). *An introduction to brain and behavior* (Vol. 3). Worth Publishers New York.

Kowler, E. (2011). Eye movements: The past 25years. *Vision Research*, *51*(13), 1457–1483. https://doi.org/10.1016/j.visres.2010.12.014

Kowler, E., & Blaser, E. (1995). The accuracy and precision of saccades to small and large targets. *Vision Research*, *35*(12), 1741–1754. https://doi.org/10.1016/0042-6989(94)00255-k

Kowler, E., & Steinman, R. M. (1980). Small saccades serve no useful purpose: Reply to a letter by r. w. ditchburn. *Vision Research*, *20*(3), 273–276. https://doi.org/10.1016/0042-6989(80)90113-3

Krauskopf, J., Cornsweet, T. N., & Riggs, L. A. (1960). Analysis of eye movements during monocular and binocular fixation. *J. Opt. Soc. Am.*, *50*(6), 572–578. https://doi.org/10.1364/JOSA.50.000572

Krieger, G., Rentschler, I., Hauske, G., Schill, K., & Zetzsche, C. (2000). Object and scene analysis by saccadic eye-movements: An investigation with higher-order statistics. *Spatial Vision*, *13*(2), 201–214. https://doi.org/10.1163/156856800741216

Kroell, L. M., & Rolfs, M. (2021). The peripheral sensitivity profile at the saccade target reshapes during saccade preparation. *Cortex*, *139*, 12–26. https://doi.org/10.1016/j.cortex.2021.02.021

Kroell, L. M., & Rolfs, M. (2022). Foveal vision anticipates defining features of eye movement targets. *eLife*, *11*. https://doi.org/10.7554/elife.78106

Kruschke, J. (2014). *Doing bayesian data analysis: A tutorial with r, jags, and stan*. Academic Press.

Kuang, X., Poletti, M., Victor, J. D., & Rucci, M. (2012). Temporal encoding of spatial information during active visual fixation. *Current Biology*, *22*(6), 510–514. https://doi.org/10.1016/j.cub.2012.01.050

Kucharsky, S., van Renswoude, D., Raijmakers, M., & Visser, I. (2021). WALD-EM: Wald accumulation for locations and durations of eye movements. *Psychological Review*, *128*(4), 667–689. https://doi.org/10.1037/rev0000292

Kümmerer, M., Wallis, T. S. A., Gatys, L. A., & Bethge, M. (2017). Understanding low- and high-level contributions to fixation prediction. *2017 Ieee International Conference on Computer Vision (iccv)*, 4799–4808. https://doi.org/10.1109/ICCV.2017.513

Kümmerer, M., & Bethge, M. (2021). *State-of-the-art in human scanpath prediction* [Preprint]. ArXiv. https://doi.org/10.48550/arXiv.2102.12239

Kümmerer, M., Theis, L., & Bethge, M. (2014). *Deep Gaze I: Boosting saliency prediction with feature maps trained on imagenet* [Preprint]. ArXiv. https://doi.org/10.48550/arXiv.1411.1045

Kümmerer, M., Wallis, T., & Bethge, M. (2014). *How close are we to understanding image-based saliency?* [Preprint]. ArXiv. https://doi.org/10.48550/arXiv.1409.7686

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

Kümmerer, M., Wallis, T. S. A., & Bethge, M. (2018). Saliency benchmarking made easy: Separating models, maps and metrics. *European Conference on Computer Vision (eccv)*. https://doi.org/10.1007/978-3-030-01270-0_47

Laloy, E., & Vrugt, J. A. (2012). High-dimensional posterior exploration of hydrologic models using multiple-try DREAM(ZS) and high-performance computing. *Water Resources Research*, *48*(1). https://doi.org/10.1029/2011wr010608

Land, M. F., & Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *352*(1358), 1231–1239. https://doi.org/10.1098/rstb.1997.0105

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, *41*(25), 3559–3565. https://doi.org/10.1016/s0042-6989(01)00102-x

Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature*, *369*(6483), 742–744. https://doi.org/10.1038/369742a0

Land, M. F., & McLeod, P. (2000). From eye movements to actions: How batsmen hit the ball. *Nature Neuroscience*, *3*(12), 1340–1345. https://doi.org/10.1038/81887

Land, M. F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*(11), 1311–1328. https://doi.org/10.1068/p2935

Land, M. F., & Tatler, B. (2009). *Looking and acting: Vision and eye movements in natural behaviour*. Oxford University Press.

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

Le Meur, O., & Baccino, T. (2012). Methods for comparing scanpaths and saliency maps: Strengths and weaknesses. *Behavior Research Methods*, *45*, 251–266. https://doi.org/10.3758/s13428-012-0226-9

Le Meur, O., & Coutrot, A. (2016). Introducing context-dependent and spatially-variant viewing biases in saccadic models. *Vision Research*, *121*, 72–84. https://doi.org/10.1016/j.visres.2016.01.005

Le Meur, O., Coutrot, A., Liu, Z., Rama, P., Le Roch, A., & Helo, A. (2017). Visual attention saccadic models learn to emulate gaze patterns from childhood to adulthood. *IEEE Transactions on Image Processing*, *26*(10), 4777–4789. https://doi.org/10.1109/tip.2017.2722238

Le Meur, O., & Liu, Z. (2015). Saccadic model of eye movements for free-viewing condition. *Vision Research*, *116*, 152–164. https://doi.org/10.1016/j.visres.2014.12.026

Li, Z. (2002). A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, *6*(1), 9–16. https://doi.org/10.1016/s1364-6613(00)01817-9

Liu, H., Xu, D., Huang, Q., Li, W., Xu, M., & Lin, S. (2013). Semantically-based human scanpath estimation with HMMs. *2013 IEEE International Conference on Computer Vision*, 3232–3239. https://doi.org/10.1109/iccv.2013.401

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(4), 565–572. https://doi.org/10.1037/0096-1523.4.4.565

Luce, R. D., & Raiffa, H. (1989). *Games and decisions: Introduction and critical survey*. Courier Corporation.

Luke, S. G., Smith, T. J., Schmidt, J., & Henderson, J. M. (2014). Dissociating temporal inhibition of return and saccadic momentum across multiple eye-movement tasks. *Journal of Vision*, *14(14):9*. https://doi.org/10.1167/14.14.9

Luke, S. G., Nuthmann, A., & Henderson, J. M. (2013). Eye movement control in scene viewing and reading: Evidence from the stimulus onset delay paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 10–15. https://doi.org/10.1037/a0030392

Maass, W., Parsons, J., Purao, S., Storey, V. C., & Woo, C. (2018). Data-driven meets theory-driven research in the era of big data: Opportunities and challenges for information systems research. *Journal of the Association for Information Systems*, 1253–1273. https://doi.org/10.17705/1jais.00526

Makowski, S., Jäger, L. A., Schwetlick, L., Trukenbrod, H. A., Engbert, R., & Scheffer, T. (2020). Discriminative viewer identification using generative models of eye gaze. *Procedia Computer Science*, *176*, 1348–1357. https://doi.org/10.1016/j.procs.2020.09.144

Malem-Shinitski, N., Opper, M., Reich, S., Schwetlick, L., Seelig, S. A., & Engbert, R. (2020). A mathematical model of local and global attention in natural scene viewing. *PLoS Computational Biology*, *16*(12), 1–21. https://doi.org/10.1371/journal.pcbi.1007880

Mannan, S. K., Wooding, D. S., & Ruddock, K. H. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision*, *10*(3), 165–188. https://doi.org/10.1163/156856896x00123

Mannan, S. K., Wooding, D. S., & Ruddock, K. H. (1997). Fixation sequences made during visual examination of briefly presented 2D images. *Spatial Vision*, *11*(2), 157–178. https://doi.org/10.1163/156856897x00177

Marino, A. C., & Mazer, J. A. (2016). Perisaccadic updating of visual representations and attentional states: Linking behavior and neurophysiology. *Frontiers in Systems Neuroscience*, *10*. https://doi.org/10.3389/fnsys.2016.00003

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. Freeman.

Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*, *5*(3), 229–240. https://doi.org/10.1038/nrn1348

Martinez-Conde, S., Macknik, S. L., Troncoso, X. G., & Dyar, T. A. (2006). Microsaccades counteract visual fading during fixation. *Neuron*, *49*(2), 297–305. https://doi.org/10.1016/j.neuron.2005.11.033

Martinez-Conde, S., Otero-Millan, J., & Macknik, S. L. (2013). The impact of microsaccades on vision: Towards a unified theory of saccadic function. *Nature Reviews Neuroscience*, *14*(2), 83–96. https://doi.org/10.1038/nrn3405

Mathôt, S., Melmi, J.-B., & Castet, E. (2015). Intrasaccadic perception triggers pupillary constriction. *PeerJ*, *3*, e1150. https://doi.org/10.7717/peerj.1150

Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin*, *81*(12), 899–917. https://doi.org/10.1037/h0037368

Matthews, R. (2000). Storks deliver babies (p = 0.008). *Teaching Statistics*, *22*(2), 36–38. https://doi.org/10.1111/1467-9639.00013

Matthis, J. S., Yates, J. L., & Hayhoe, M. M. (2018). Gaze and the control of foot placement when walking in natural terrain. *Current Biology*, *28*(8), 1224–1233. https://doi.org/10.1016/j.cub.2018.03.008

McCamy, M., Macknik, S. L., & Martinez-Conde, S. (2014). Different fixational eye movements mediate the prevention and the reversal of visual fading. *Journal of Physiology*, *592*(19), 4381–4394. https://doi.org/10.1113/jphysiol.2014.279059

McCamy, M., Otero-Millan, J., Macknik, S., Yang, Y., Troncoso, X., Baer, S., Crook, S., & Martinez-Conde, S. (2012). Microsaccadic efficacy and contribution to foveal and peripheral vision. *Journal of Vision*, *12(9):15*. https://doi.org/10.1167/12.9.1015

Meese, T. S., Georgeson, M. A., & Baker, D. H. (2006). Binocular contrast vision at and above threshold. *Journal of Vision*, *6(11):7*. https://doi.org/10.1167/6.11.7

Mergenthaler, K., & Engbert, R. (2010). Microsaccades are different from saccades in scene perception. *Experimental Brain Research*, *203*(4), 753–757. https://doi.org/10.1007/s00221-010-2272-9

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, *21*(6), 1087–1092. https://doi.org/10.1063/1.1699114

Metzler, R., & Klafter, J. (2000). The random walk's guide to anomalous diffusion: A fractional dynamics approach. *Physics Reports*, *339*(1), 1–77. https://doi.org/10.1016/s0370-1573(00)00070-3

Mills, M., Hollingworth, A., van der Stigchel, S., Hoffman, L., & Dodd, M. D. (2011). Examining the influence of task set on eye movements and fixations. *Journal of Vision*, *11(8):17*. https://doi.org/10.1167/11.8.17

Mirpour, K., Bolandnazar, Z., & Bisley, J. W. (2019). Neurons in FEF keep track of items that have been previously fixated in free viewing visual search. *Journal of Neuroscience*, *39*(11), 2114–2124. https://doi.org/10.1523/jneurosci.1767-18.2018

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, *47*(1), 90–100. https://doi.org/10.1016/s0022-2496(02)00028-7

Nachmias, J. (1959). Two-dimensional motion of the retinal image during monocular fixation. *Journal of the Optical Society of America*, *49*(9), 901. https://doi.org/10.1364/josa.49.000901

Næss, S., Halnes, G., Hagen, E., Hagler, D. J., Dale, A. M., Einevoll, G. T., & Ness, T. V. (2021). Biophysically detailed forward modeling of the neural origin of EEG and MEG signals. *NeuroImage*, *225*, 117467. https://doi.org/10.1016/j.neuroimage.2020.117467

Newell, A., Simon, H. A., et al. (1972). *Human problem solving* (Vol. 104). Prentice-hall Englewood Cliffs.

Nielsen, M. (2015). *Neural networks and deep learning*. Determination Press.

Noorani, I., & Carpenter, R. H. S. (2016). The LATER model of reaction time and decision. *Neuroscience & Biobehavioral Reviews*, *64*, 229–251. https://doi.org/10.1016/j.neubiorev.2016.02.018

Noton, D., & Stark, L. (1971a). Scanpaths in eye movements during pattern perception. *Science*, *171*(3968), 308–311. https://doi.org/10.1126/science.171.3968.308

Noton, D., & Stark, L. (1971b). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research*, *11*(9), 929–942. https://doi.org/10.1016/0042-6989(71)90213-6

Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision*, *10(8):20*. https://doi.org/10.1167/10.8.20

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology/Revue Canadienne De Psychologie*, *46*(3), 461–488. https://doi.org/10.1037/h0084327

Otero-Millan, J., Macknik, S. L., Langston, R. E., & Martinez-Conde, S. (2013). An oculomotor continuum from exploration to fixation. *Proceedings of the National Academy of Sciences*, *110*(15), 6175–6180. https://doi.org/10.1073/pnas.1222715110

Otero-Millan, J., Macknik, S. L., Serra, A., Leigh, R. J., & Martinez-Conde, S. (2011). Triggering mechanisms in microsaccade and saccade generation: A novel proposal. *Annals of the New York Academy of Sciences*, *1233*(1), 107–116. https://doi.org/10.1111/j.1749-6632.2011.06177.x

Over, E. A. B., Hooge, I. T. C., Vlaskamp, B. N. S., & Erkelens, C. J. (2007). Coarse-to-fine eye movement strategy in visual search. *Vision Research*, *47*(17), 2272–2280. https://doi.org/10.1016/j.visres.2007.05.002

Packer, O., & Williams, D. R. (1992). Blurring by fixational eye movements. *Vision Research*, *32*(10), 1931–1939. https://doi.org/10.1016/0042-6989(92)90052-k

Pan, J., McGuinness, K., Sayrol, E., O'Connor, N., & Giro-i-Nieto, X. (2016). *Shallow and deep convolutional networks for saliency prediction* [Preprint]. ArXiv. https://doi.org/10.48550/arXiv.1603.00845

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, *42*(1), 107–123. https://doi.org/10.1016/s0042-6989(01)00250-4

Parkhurst, D., & Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision*, *16*(2), 125–154. https://doi.org/10.1163/15685680360511645

Pedziwiatr, M. A., Kümmerer, M., Wallis, T. S. A., Bethge, M., & Teufel, C. (2021a). Meaning maps and saliency models based on deep convolutional neural networks are insensitive to image meaning when predicting human fixations. *Cognition*, *206*, 104465. https://doi.org/10.1016/j.cognition.2020.104465

Pedziwiatr, M. A., Kümmerer, M., Wallis, T. S. A., Bethge, M., & Teufel, C. (2021b). There is no evidence that meaning maps capture semantic information relevant to gaze guidance: Reply to henderson, hayes, peacock, and rehrig (2021). *Cognition*, *214*, 104741. https://doi.org/10.1016/j.cognition.2021.104741

Pelz, J. B., & Canosa, R. (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research*, *41*(25-26), 3587–3596. https://doi.org/10.1016/s0042-6989(01)00245-0

Peterson, M. F., & Eckstein, M. P. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences*, *109*(48), E3314–E3323. https://doi.org/10.1073/pnas.1214269109

Peterson, M. F., & Eckstein, M. P. (2013). Individual differences in eye movements during face identification reflect observer-specific optimal points of fixation. *Psychological Science*, *24*(7), 1216–1225. https://doi.org/10.1177/0956797612471684

Pitkow, X., Sompolinsky, H., & Meister, M. (2007). A neural computation for visual acuity in the presence of eye movements (D. Burr, Ed.). *PLoS Biology*, *5*(12), e331. https://doi.org/10.1371/journal.pbio.0050331

Poletti, M., Listorti, C., & Rucci, M. (2013). Microscopic eye movements compensatefor nonhomogeneous vision within the fovea. *Current Biology*, *23*(17), 1691–1695. https://doi.org/10.1016/j.cub.2013.07.007

Poletti, M., & Rucci, M. (2016). A compact field guide to the study of microsaccades: Challenges and functions. *Vision Research*, *118*, 83–97. https://doi.org/10.1016/j.visres.2015.01.018

Polyak, S. L. (1941). *The retina.* Chicago: University of Chicago Press.

Port, R. F., & van Gelder, T. (Eds.). (1995). *Mind as motion.* MIT Press.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Posner, M. I., & Cohen, Y. (1984). Attention and performance x: Control of language processes. In H. Bouma & D. G. Bouwhuis (Eds.). Lawrence Erlbaum.

Posner, M. I., Rafal, R. D., Choate, L. S., & Vaughan, J. (1985). Inhibition of return: Neural basis and function. *Cognitive Neuropsychology*, *2*(3), 211–228. https://doi.org/10.1080/02643298508252866

Poynter, W., Barber, M., Inman, J., & Wiggins, C. (2013). Individuals exhibit idiosyncratic eye-movement behavior profiles across tasks. *Vision Research*, *89*, 32–38. https://doi.org/10.1016/j.visres.2013.07.002

Privitera, C. M., & Stark, L. W. (2000). Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(9), 970–982. https://doi.org/10.1109/34.877520

Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W. C., LaMantia, A.-S., & White, L. E. (1997). *Neuroscience.* Sunderland: Sinauer Associates. https://doi.org/10.4249/scholarpedia.7204

R Core Team. (2019). *R: A language and environment for statistical computing* (Version 3.6.2) [Computer Scoftware]. Vienna, Austria. https://www.R-project.org/

Rabe, M. M., Chandra, J., Krügel, A., Seelig, S. A., Vasishth, S., & Engbert, R. (2021). A Bayesian approach to dynamical modeling of eye-movement control in reading of normal, mirrored, and scrambled texts. *Psychological Review*, *128*(5), 803–823. Advance online publication. https://doi.org/10.1037/rev0000268

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420

Ratliff, F., & Riggs, L. A. (1950). Involuntary motions of the eye during monocular fixation. *Journal of Experimental Psychology*, *40*(6), 687–701. https://doi.org/10.1037/h0057754

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Rayner, K. (1995). Eye movements and cognitive processes in reading, visual search, and scene perception. In *Studies in visual information processing* (pp. 3–22). Elsevier. https://doi.org/10.1016/s0926-907x(05)80003-0

Rayner, K., Smith, T. J., Malcolm, G. L., & Henderson, J. M. (2009). Eye movements and visual encoding during scene perception. *Psychological Science*, *20*(1), 6–10. https://doi.org/10.1111/j.1467-9280.2008.02243.x

Reddi, B. A. J., & Carpenter, R. H. S. (2000). The influence of urgency on decision time. *Nature Neuroscience*, *3*(8), 827–830. https://doi.org/10.1038/77739

Reich, S., & Cotter, C. (2015). *Probabilistic forecasting and bayesian data assimilation*. Cambridge University Press.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*(1), 125–157. https://doi.org/10.1037/0033-295x.105.1.125

Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network: Computation in Neural Systems*, *10*(4), 341–350. https://doi.org/10.1088/0954-898X/10/4/304

Reingold, E. M., & Sheridan, H. (2011). Eye movements and visual expertise in chess and medicine. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *Oxford handbook on eye movements* (pp. 523–550). Oxford University Press. https://eprints.soton.ac.uk/367506/

Ritter, F. E., Tehranchi, F., & Oury, J. D. (2018). ACT-r: A cognitive architecture for modeling cognition. *WIREs Cognitive Science*, *10*(3), e1488. https://doi.org/10.1002/wcs.1488

Roberts, J. A., Wallis, G., & Breakspear, M. (2013). Fixational eye movements during viewing of dynamic natural scenes. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00797

Rolfs, M., Kliegl, R., & Engbert, R. (2008). Toward a model of microsaccade generation: The case of microsaccadic inhibition. *Journal of Vision*, *8(11):5*. https://doi.org/10.1167/8.11.5

Rolfs, M., Jonikaitis, D., Deubel, H., & Cavanagh, P. (2011). Predictive remapping of attention across eye movements. *Nature Neuroscience*, *14*(2), 252–256. https://doi.org/10.1038/nn.2711

Roth, N., Rolfs, M., & Obermayer, K. (2022). Scanpath prediction in dynamic real-world scenes based on object-based selection. *Journal of Vision*, *22(14):4217*. https://doi.org/10.1167/jov.22.14.4217

Rothkegel, L. O. M., Schütt, H., Trukenbrod, H. A., Wichmann, F., & Engbert, R. (2019). *Potsdam scene viewing corpus* [Data set]. Open Science Framework. https://doi.org/10.17605/OSF.IO/N3BYQ

Rothkegel, L. O. M., Schütt, H. H., Trukenbrod, H. A., Wichmann, F. A., & Engbert, R. (2019). Searchers adjust their eye-movement dynamics to target characteristics in natural scenes. *Scientific Reports*, *9*, 1635. https://doi.org/10.1038/s41598-018-37548-w

Rothkegel, L. O. M., Schütt, H. H., Trukenbrod, H. A., Wichmann, F. A., & Engbert, R. (2018). *Searchers adjust their eye movement dynamics to the target characteristics in natural scenes* [Preprint]. ArXiv. https://doi.org/10.48550/arXiv.1802.04069

Rothkegel, L. O. M., Trukenbrod, H. A., Schütt, H. H., Wichmann, F. A., & Engbert, R. (2016). Influence of initial fixation position in scene viewing. *Vision Research*, *129*, 33–49. https://doi.org/10.1016/j.visres.2016.09.012

Rothkegel, L. O. M., Trukenbrod, H. A., Schütt, H. H., Wichmann, F. A., & Engbert, R. (2017). Temporal evolution of the central fixation bias in scene viewing. *Journal of Vision*, *17(13):3*. https://doi.org/10.1167/17.13.3

Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *Journal of Vision*, *7(14):16*. https://doi.org/10.1167/7.14.16

Rucci, M., Iovin, R., Poletti, M., & Santini, F. (2007). Miniature eye movements enhance fine spatial detail. *Nature*, *447*(7146), 852–855. https://doi.org/10.1038/nature05866

Rucci, M., & Poletti, M. (2015). Control and functions of fixational eye movements. *Annual Review of Vision Science*, *1*(1), 499–518. https://doi.org/10.1146/annurev-vision-082114-035742

Rucci, M., & Victor, J. D. (2015). The unsteady eye: An information-processing stage, not a bug. *Trends in Neurosciences*, *38*(4), 195–206. https://doi.org/10.1016/j.tins.2015.01.005

Schad, D. J., Betancourt, M., & Vasishth, S. (2021). Toward a principled Bayesian workflow in cognitive science. *Psychological Methods*, *26*(1), 103–126. https://doi.org/10.1037/met0000275

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038. https://doi.org/10.1016/j.jml.2019.104038

Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky, C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., & Ajoy, Y. G. A. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution*, *11*(9), 1141–1152. https://doi.org/10.1111/2041-210X.13434

Schmittwilken, L., & Maertens, M. (2022). Fixational eye movements enable robust edge detection. *Journal of Vision*, *22(8):5*. https://doi.org/10.1167/jov.22.8.5

Schneider, W. X., & Deubel, H. (1995). Visual attention and saccadic eye movements: Evidence for obligatory and selective spatial coupling. In J. M. Findlay, R. Walker, & R. W. Kentridge (Eds.), *Eye movement research: Mechanisms, processes and applications* (pp. 317–324). Amsterdam: Elsevier. https://doi.org/10.1016/s0926-907x(05)80027-3

Schütt, H. H., Rothkegel, L. O. M., Trukenbrod, H. A., Engbert, R., & Wichmann, F. A. (2019). Disentangling bottom-up versus top-down and low-level versus high-level influences on eye movements over time. *Journal of Vision*, *19(3):1*. https://doi.org/10.1167/19.3.1

Schütt, H. H., Rothkegel, L. O. M., Trukenbrod, H. A., Reich, S., Wichmann, F. A., & Engbert, R. (2017). Likelihood-based parameter estimation and comparison of dynamical cognitive models. *Psychological Review*, *124*(4), 505–524. https://doi.org/10.1037/rev0000068

Schwetlick, L., Backhaus, D., Brunken, R., & Engbert, R. (2022). *The effect of illumination-level on measurement stability using an eyelink1000 eye tracker* [Preregistration]. https://doi.org/10.17605/OSF.IO/3GUK4

Schwetlick, L., Backhaus, D., & Engbert, R. (2020). *Modelling advanced natural tasks using scenewalk* [Preregistration]. OSF. https://osf.io/dsyt2/

Schwetlick, L., Backhaus, D., & Engbert, R. (2022a). A dynamical scan-path model for task-dependence during scene viewing. *Psychological Review*, *130*(3), 807–8. https://doi.org/10.1037/rev0000379

Schwetlick, L., Backhaus, D., & Engbert, R. (2022b). Modeling task-dependency of eye movement during scene viewing. In V. McGowan, A. Pagán, K. B. Paterson, D. Souto, & R. Groner (Eds.), *Book of abstracts of the 21st european conference on eye movements*. Journal of Eye Movement Research, 15 *(5)*. https://doi.org/10.16910/jemr.15.5.1

Schwetlick, L., Backhaus, D., Trukenbrod, H. A., & Engbert, R. (2020). *"Memory": Image familiarity and eye movement* [Data set]. Open Science Framework. https://doi.org/10.17605/OSF.IO/E7FVP

Schwetlick, L., Kümmerer, M., Bethge, M., & Engbert, R. (2022). *Potsdam dataset for eye movement on natural scenes (potsdam daemons)* [Preregistration]. https://doi.org/10.17605/OSF.IO/BDXGS

Schwetlick, L., Kümmerer, M., Engbert, R., & Bethge, M. (2022). DeepGaze vs SceneWalk: What can DNNs and biological scan path models teach each other? *Journal of Vision*, *22(14):3986*. https://doi.org/10.1167/jov.22.14.3986

Schwetlick, L., Reich, S., & Engbert, R. (2023). *Bayesian dynamical modeling of fixational eye movements* [Preprint]. ArXiv. https://doi.org/10.48550/arXiv.2303.11941

Schwetlick, L., Rothkegel, L. O. M., & Engbert, R. (2019). Adding neurally-inspired mechanisms to the SceneWalk model improves scan path predictions for natural images. *2019 Conference on Cognitive Computational Neuroscience*. https://doi.org/10.32470/ccn.2019.1206-0

Schwetlick, L., Rothkegel, L. O. M., & Engbert, R. (2020). Peri-saccadic attention drives saccade statistics in scene viewing. *Journal of Vision*, *20(11):700*. https://doi.org/10.1167/jov.20.11.700

Schwetlick, L., Rothkegel, L. O. M., Trukenbrod, H. A., & Engbert, R. (2017). Central fixation bias: The role of sudden image onset and early gist extraction. *European Conference on Visual Perception 2017*. https://doi.org/10.16910/jemr.10.6.1

Schwetlick, L., Rothkegel, L. O. M., Trukenbrod, H. A., & Engbert, R. (2020a). *Modeling the effects of perisaccadic attention on gaze statistics during scene viewing* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/zcbny

Schwetlick, L., Rothkegel, L. O. M., Trukenbrod, H. A., & Engbert, R. (2020b). Modeling the effects of perisaccadic attention on gaze statistics during scene viewing. *Communications Biology*, *3*(727), 1–11. https://doi.org/10.1038/s42003-020-01429-8

Schwetlick, L., Trukenbrod, H. A., & Engbert, R. (2018). The influence of visual long term memory on eye movements during scene viewing. *European Conference on Visual Perception 2018*.

Schwetlick, L., Trukenbrod, H. A., & Engbert, R. (2019). The effect of visual long-term memory on eye movements over time. *Journal of Vision*, *19(10):149a*. https://doi.org/10.1167/19.10.149a

Seelig, S. A., Rabe, M. M., Malem-Shinitski, N., Risse, S., Reich, S., & Engbert, R. (2020). Bayesian parameter estimation for the SWIFT model of eye-movement control during reading. *Journal of Mathematical Psychology*, *95*, 102313. https://doi.org/10.1016/j.jmp.2019.102313

SensoMotoric Instruments. (2016). *iViewETG user guide*. SensoMotoric Instruments.

Shao, X., Luo, Y., Zhu, D., Li, S., Itti, L., & Lu, J. (2017). Scanpath prediction based on high-level features and memory bias. In *International conference on neural information processing* (pp. 3–13). Springer International Publishing. https://doi.org/10.1007/978-3-319-70090-8_1

Shelchkova, N., Tang, C., & Poletti, M. (2019). Task-driven visual exploration at the foveal scale. *Proceedings of the National Academy of Sciences*, *116*(12), 5811–5818. https://doi.org/10.1073/pnas.1812222116

Shockley, E. M., Vrugt, J. A., & Lopez, C. F. (2018). PyDREAM: high-dimensional parameter inference for biological models in python. *Bioinformatics*, *34*(4), 695–697. https://doi.org/10.1093/bioinformatics/btx626

Shulman, G. L., Remington, R. W., & Mclean, J. P. (1979). Moving attention through visual space. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(3), 522–526. https://doi.org/10.1037/0096-1523.5.3.522

Shurygina, O., Pooresmaeili, A., & Rolfs, M. (2021). Pre-saccadic attention spreads to stimuli forming a perceptual group with the saccade target. *Cortex*, *140*, 179–198. https://doi.org/10.1016/j.cortex.2021.03.020

Simonyan, K., & Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition* [Preprint]. ArXiv. https://doi.org/10.48550/arXiv.1409.1556

Sinn, P., & Engbert, R. (2016). Small saccades versus microsaccades: Experimental distinction and model-based unification. *Vision Research*, *118*, 132–143. https://doi.org/10.1016/j.visres.2015.05.012

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, *27*(3), 161–168. https://doi.org/10.1016/j.tins.2004.01.006

Smith, T. J., & Henderson, J. M. (2009). Facilitation of return during scene viewing. *Visual Cognition*, *17*(6-7), 1083–1108. https://doi.org/10.1080/13506280802678557

Sparks, D. L. (2002). The brainstem control of saccadic eye movements. *Nature Reviews Neuroscience*, *3*(12), 952–964. https://doi.org/https://doi.org/10.1038/nrn986

Spauschus, A., Marsden, J., Halliday, D. M., Rosenberg, J. R., & Brown, P. (1999). The origin of ocular microtremor in man. *Experimental Brain Research*, *126*(4), 556–562. https://doi.org/10.1007/s002210050764

Spencer, R. F., & Porter, J. D. (2006). Biological organization of the extraocular muscles. In *Progress in brain research* (pp. 43–80). Elsevier. https://doi.org/10.1016/s0079-6123(05)51002-1

Storn, R., & Price, K. (1997). *Journal of Global Optimization*, *11*(4), 341–359. https://doi.org/10.1023/a:1008202821328

Strasburger, H., Rentschler, I., & Juttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision*, *11(5):13*. https://doi.org/10.1167/11.5.13

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

Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research*, *45*(5), 643–659. https://doi.org/10.1016/j.visres.2004.09.017

Tatler, B. W., Baddeley, R. J., & Vincent, B. T. (2006). The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task. *Vision Research*, *46*(12), 1857–1862. https://doi.org/10.1016/j.visres.2005.12.005

Tatler, B. W., Brockmole, J. R., & Carpenter, R. H. S. (2017). LATEST: A model of saccadic decisions in space and time. *Psychological Review*, *124*(3), 267–300. https://doi.org/10.1037/rev0000054

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, *11(5):5*. https://doi.org/10.1167/11.5.5

Tatler, B. W., Vincent, B. T., et al. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, *2*(2). https://doi.org/10.16910/jemr.2.2.5

Tatler, B. W., & Vincent, B. T. (2009). The prominence of behavioural biases in eye guidance. *Visual Cognition*, *17*(6-7), 1029–1054. https://doi.org/10.1080/13506280902764539

Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: Capture of the eyes by new objects. *Psychological Science*, *9*(5), 379–385. https://doi.org/10.1111/1467-9280.00071

Thiel, M., Romano, M. C., Kurths, J., Rolfs, M., & Kliegl, R. (2006). Twin surrogates to test for complex synchronisation. *Europhysics Letters (EPL)*, *75*(4), 535–541. https://doi.org/10.1209/epl/i2006-10147-0

Tian, X., Yoshida, M., & Hafed, Z. M. (2016). A microsaccadic account of attentional capture and inhibition of return in posner cueing. *Frontiers in Systems Neuroscience*, *10*. https://doi.org/10.3389/fnsys.2016.00023

Tian, X., Yoshida, M., & Hafed, Z. M. (2018). Dynamics of fixational eye position and microsaccades during spatial cueing: The case of express microsaccades. *Journal of Neurophysiology*, *119*(5), 1962–1980. https://doi.org/10.1152/jn.00752.2017

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, *113*(4), 766–786. https://doi.org/10.1037/0033-295x.113.4.766

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1093/acprof:osobl/9780199734337.003.0011

Trukenbrod, H. A., Barthelmé, S., Wichmann, F. A., & Engbert, R. (2019). Spatial statistics for gaze patterns in scene viewing: Effects of repeated viewing. *Journal of Vision*, *19(6):5*. https://doi.org/10.1167/19.6.5

Trukenbrod, H. A., & Engbert, R. (2014). Icat: A computational model for the adaptive control of fixation durations. *Psychonomic Bulletin & Review*, *21*(4), 907–934. https://doi.org/10.3758/s13423-013-0575-0

Trukenbrod, H. A., Schwetlick, L., & Engbert, R. (2020). *Spatial statistics for gaze patterns of repeated viewing in scene perception* [Data Set]. Open Science Framework. https://doi.org/10.17605/OSF.IO/ME2SH

Tsotsos, J. K. (1990). A complexity level analysis of vision. *Behavioral and Brain Sciences*, *13*(3), 423–445. https://doi.org/10.1017/S0140525X00079577

Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence*, *78*(1-2), 507–545. https://doi.org/10.1016/0004-3702(95)00025-9

Underwood, G., Chapman, P., Brocklehurst, N., Underwood, J., & Crundall, D. (2003). Visual attention while driving: Sequences of eye fixations made by experienced and novice drivers. *Ergonomics*, *46*(6), 629–646. https://doi.org/10.1080/0014013031000090116

Underwood, G., Crundall, D., & Chapman, P. (2007). Driving. In *Handbook of applied cognition* (pp. 391–414). John Wiley & Sons Ltd. https://doi.org/10.1002/9780470713181.ch15

Valsecchi, M., Betta, E., & Turatto, M. (2006). Visual oddballs induce prolonged microsaccadic inhibition. *Experimental Brain Research*, *177*(2), 196–208. https://doi.org/10.1007/s00221-006-0665-6

Van Gelder, T., & Robert, P. (1995). Mind as motion. In R. F. Port & T. van Gelder (Eds.). MIT Press.

van Renswoude, D. R., van den Berg, L., Raijmakers, M. E. J., & Visser, I. (2019). Infants' center bias in free viewing of real-world scenes. *Vision Research*, *154*, 44–53. https://doi.org/10.1016/j.visres.2018.10.003

Vansteenkiste, P., Van Hamme, D., Veelaert, P., Philippaerts, R., Cardon, G., & Lenoir, M. (2014). Cycling around a curve: The effect of cycling speed on steering and gaze behavior. *PLoS ONE*, *9*(7), e102792. https://doi.org/10.1371/journal.pone.0102792

von Wartburg, R., Wurtz, P., Pflugshaupt, T., Nyffeler, T., Lüthi, M., & Müri, R. M. (2007). Size matters: Saccades during scene perception. *Perception*, *36*(3), 355–365. https://doi.org/10.1068/p5552

Vrugt, J. A., & Braak, C. J. F. T. (2011). DREAM: An adaptive markov chain monte carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problems. *Hydrology and Earth System Sciences*, *15*(12), 3701–3713. https://doi.org/10.5194/hess-15-3701-2011

Wilming, N., Harst, S., Schmidt, N., & König, P. (2013). Saccadic momentum and facilitation of return saccades contribute to an optimal foraging strategy. *PLoS Computational Biology*, *9*(1), e1002871. https://doi.org/10.1371/journal.pcbi.1002871

Winterson, B. J., & Collewijn, H. (1976). Microsaccades during finely guided visuomotor tasks. *Vision Research*, *16*(12), 1387–1390. https://doi.org/10.1016/0042-6989(76)90156-5

Wolfe, J. (2015). The handbook of attention. In J. Fawcett, E. Risko, & A. Kingstone (Eds.). MIT Press. https://books.google.de/books?id=zswhCwAAQBAJ

Yantis, S., & Abrams, R. A. (2014). *Sensation and perception*. Worth Publishers New York.
Yarbus, A. L. (1967). *Eye movements and vision*. Plenum Press. https://doi.org/10.1007/978-1-4899-5379-7
YutaItoh. (2016). *3d eye tracker* [Computer software]. GitHub. https://github.com/YutaItoh/3D-Eye-Tracker
Zanca, D., Melacci, S., & Gori, M. (2020). Gravitational laws of focus of attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(12), 2983–2995. https://doi.org/10.1109/tpami.2019.2920636
Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review*, *115*(4), 787–835. https://doi.org/10.1037/a0013118
Zhaoping, L. (2014). *Understanding vision: Theory, models, and data*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199564668.001.0001
Zhou, Y., & Yu, Y. (2021). Human visual search follows a suboptimal Bayesian strategy revealed by a spatiotemporal computational model and experiment. *Communications Biology*, *4*(1), 1–16. https://doi.org/10.1038/s42003-020-01485-0
Zuber, B. L., Stark, L., & Cook, G. (1965). *Science*, *150*(3702), 1459–1460. https://doi.org/10.1126/science.150.3702.1459

A Appendix for Paper 1

A.1 Supplementary methods

1.1.1 Example images used in the experiment

The images used as stimuli in this study represent a subset of the Potsdam Corpus on Spatial Frequency Search in Natural Scenes (Rothkegel, Schütt, et al., 2019). Examples are given in Figure A.1.

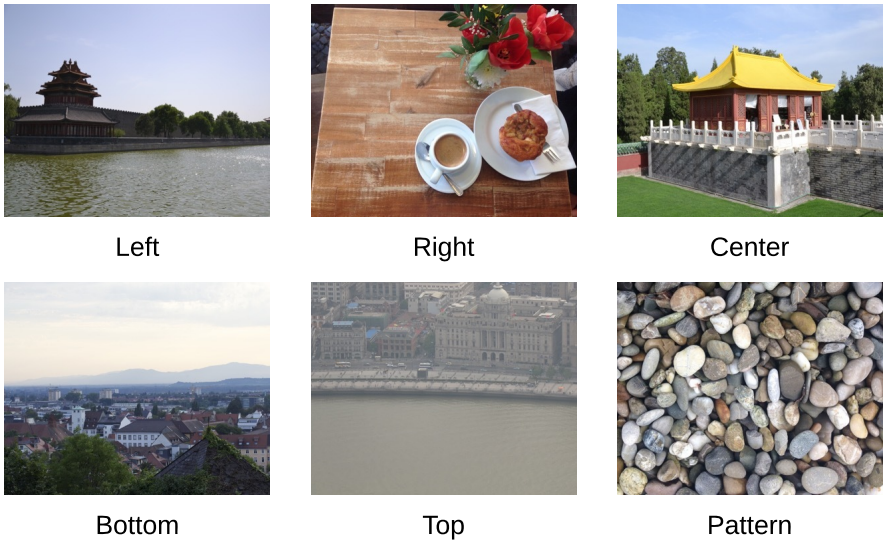

Figure A.1 Six sample images from the scene viewing corpus. The first images are representative for categories of left, right, central, bottom, and top focus. The bottom right image provides an example for a natural pattern.

1.1.2 Model phases

In Figure A.2 we show an alternative visualization of the temporal progression in the model.

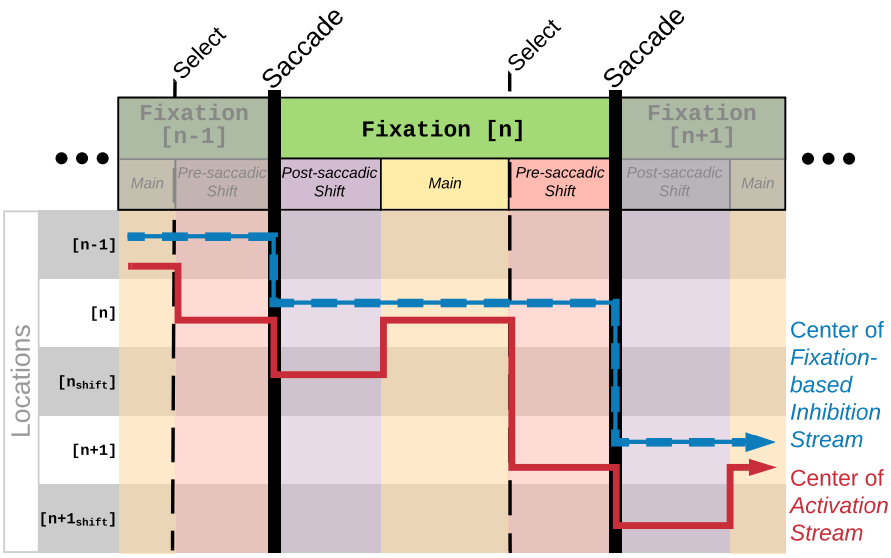

Figure A.2 Fixation-phases of the activation and inhibition streams in the extended SceneWalk Model. The blue line shows the location around which the fixation-based inhibition stream's Gaussian aperture is centered. The red line represents the center of the activation stream. During each fixation, the model implements three phases. While the inhibition stream remains on the fixated location, the center of the activation stream shifts around the time of each saccade.

1.1.3 Fixed model parameters

In addition to the results on estimated model parameters we report the fixed model parameters in Table A.1.

A.2 Supplementary results

1.2.1 Results on systematic tendencies

In addition to the measures of scan path statistics reported in the *Results* we also investigated model performance with respect to two further statistics. First, as a measure of how fixations spread over an image over time we investigated the *mean lag distance*, defined as the distance between two fixations, separated by x other fixations (Fig. A.3A). Empirical data indicate that the distance between fixation n and $n + x$ separate quickly for 3 to 5 fixations before reaching peak distance and returning to chance-level distance. We interpret this overshoot-type behavior as an indication of inhibitory tagging as one of the key driving mechanisms during scene exploration. The general tendency is present in the baseline model. While the extended model improves the fit to experimental data, the overshoot in the distance is not present. From this result, we might conclude that the inhibitor component is currently too weak in both mathematical models.

Second, an important systematic bias in eye movements is the central fixation ten-

Parameter	Baseline SceneWalk	Extended Model
ω_A/ω_F	10	10
CF	0.3	0.3
τ_{pre}	–	0.05
τ_{post}	–	0.1
ν	–	2
σ_{post}	–	2
ω_{CB}	–	1.5
σ_{CBx}	–	4
σ_{CBy}	–	3
ω_A/ω_{FoR}	–	10

Table A.1 Fixed model parameters for baseline and extended model.

dency. Specifically, the first fixation in a scan path tends to be closer to center of the scene than subsequent fixations (Bindemann, 2010; Rothkegel et al., 2017; Tatler, 2007; van Renswoude et al., 2019). Figure A.3C shows the distance to the image center over fixations. We added a center bias to the model by initializing the model in the attention stream with a centered Gaussian activation map (Rothkegel et al., 2017). The characteristic dip on the second fixation (i.e., the first freely chosen fixation) is reproduced exactly by the new model.

Furthermore, the central fixation tendency is stronger when the initial saccade latency was shorter than on average (Rothkegel et al., 2017). Here we analyzed the dependence of the central fixation bias on the initial saccade latency (Figure A.3B). Note that this analysis includes the full data set (i.e., the analysis is not limited to the test data) to produce more stable results. As a result, a larger latency before the first saccade is systematically related to a reduced central fixation bias compared to shorter fixations. The changes to the model improve the dynamic dependency of this measure.

1.2.2 Individual differences

Using separate by-participant estimates, it is possible to examine how well the fitted model parameters capture inter-individual differences. For this analysis we compared the subject-specific experimental data to data simulated with the subject's set of parameters. We expect that subjects with high expression of a statistic should also show a higher degree of that statistic in simulated data. the specific model parameters fitted for each participant show good agreement between experimental and simulated data, and explain inter-individual differences, which we quantified by the corresponding correlation coefficient.

In Figure 3 we show the correlation of mean saccade amplitude in experimental and simulated data. Correspondingly, Figure A.4A shows the correlations for the mean

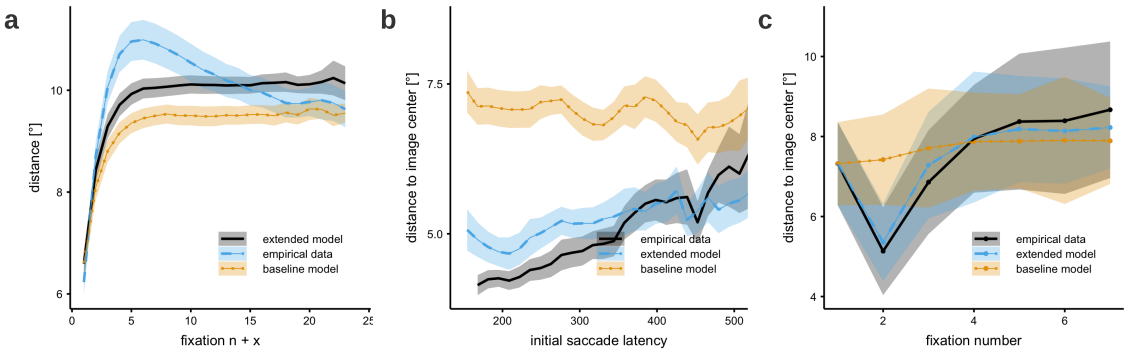

Figure A.3 Model performance on additional statistics. **(a)** Mean distance between a fixation and subsequent fixations in experimental and simulated data. **(b)** Modulation of the the first fixation's mean distance to the image center by initial saccade latency. **(c)** The distance to the image center of each fixation in the sequence. The empirical tendency to move close to the center at fixation 2 is well-replicated by the extended model.

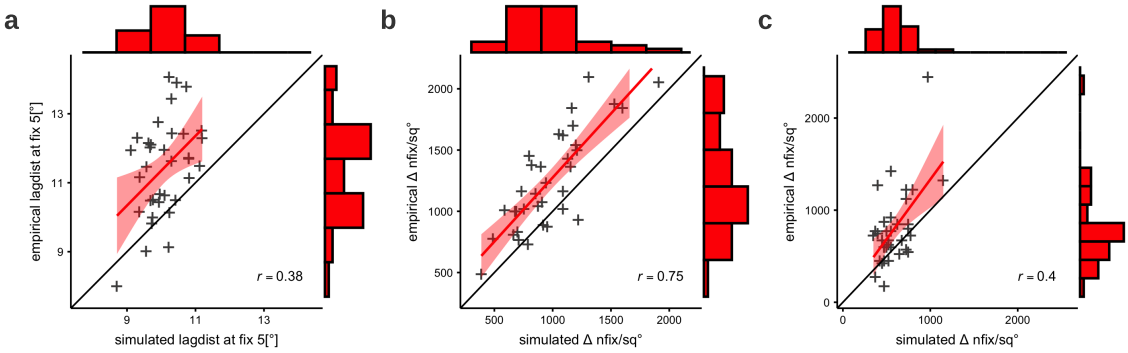

Figure A.4 Correlations between experimental and simulated data across participants. **(a)** Mean lag distance at fixation 5. **(b)** Number of fixations per square degree that land within the area expected to contain forward saccades. **(c)** Number of fixations per square degree that land in the area expected to contain return saccades.

lag distance metric. We chose the distance at fixation 5 as our correlation measure, which on average is the point of peak distance in the experimental data. Both saccade amplitude and inhibition show good agreement between experimental and simulated data, suggesting that the model parameters capture the inter-individual differences adequately. The correlation is particularly interesting given that the mean lag distance peak is not strongly present in the simulated data.

Mean saccade amplitude (Fig. 3C) is closely related to the activation stream and lag distance to the inhibition stream. Thus, as the two most fundamental mechanisms in the model, the fact that they are well-represented in the individual model fits lends support to our model.

In Figure A.4 B and C we show correlations between empirical and simulated data concerning the amount of forward and return saccades. These metrics represent the two most important new model components. The values we compare are the number

of fixations per square degree that fall within a defined window for each peak. While there is more variation between subjects in the case of forward saccades, the magnitude of each tendency is well-represented by the model fit. The positive correlation of these measures shows that the model extension capture the relevant aspects in the data and account for inter-individual differences.

In general it is important to note that interindividual differences play a large role in eye movement behavior, particularly also concerning different participant groups such as children and/or patients (Helo et al., 2014; Le Meur et al., 2017). In future work our new by-subject estimation procedure could be applied to corresponding data.

1.2.3 Parameter recovery analysis

As an indicator for the reliability of our numerical simulations and, in particular, of our statistical inference, we investigated the parameter inference on simulated data with known parameters. We simulated data using parameter point estimates for one representative participant. Simulated data were then fed back to parameter inference, as outlined above (see *Methods*). In Figure A.5 we report the posteriors over parameters in comparison to the true numerical values (χ was not included in this analysis). Thus, the combination of model simulations and the estimation procedures can recover values from the experimental data reliably. These analyses strengthen the credibility, stability, and robustness of our mathematical modeling approach.

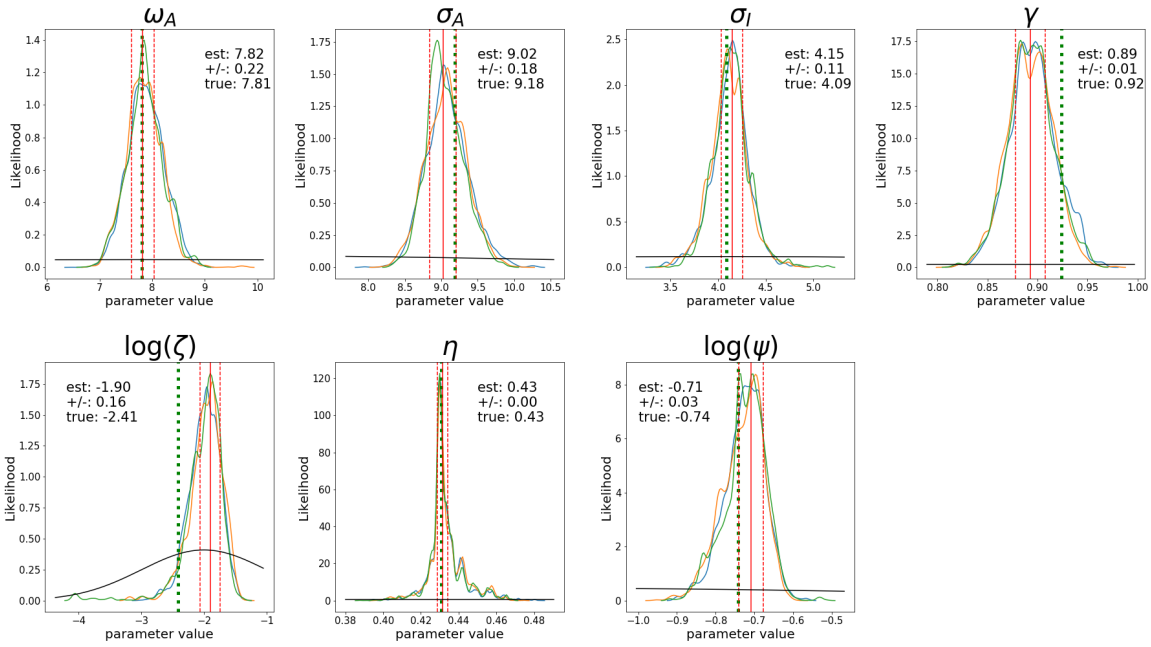

Figure A.5 Model parameter recovery analysis. Panels report posteriors for estimated model parameters obtained from simulated data. The three curves represent three estimation chains of the DREAM algorithm. The green, dotted line is the parameter value from which data was generated. The red lines show the recovered parameter values: the maximum posteriori value (solid line) and the 50% credibility interval (dotted line). Black lines show the prior.

1.2.4 Detailed results on parameter estimation

Based on the statistical methods described in the previous section, we obtained the parameter point estimates reported in Table 2 averaged across participants (see Tables A.2 and A.3).

As the parameter estimation was conducted in a fully Bayesian framework, we have also access to the full posterior likelihood distribution of each parameter. In Fig. A.6 we show the marginal posterior distributions of each parameter of the model for all subjects. The marginal posterior distributions can serve as a tool to understand how well the parameters constrain the data.

For each parameter, we link the interpretation of the marginal posteriors to the function of the corresponding parameter in the model. The speed of the decay of the attention stream ω_A controls the duration of the memory of the process for allocation to past target locations. The numerical value $\omega_A = 10.12$ is indicates a half life of 70 ms of the previous map's influence (i.e., $\exp(-10.12 \cdot 0.070) = 0.49$). The Gaussian of the attention and inhibition streams are specified by parameters $\sigma_A = 7.3$ and $\sigma_F = 6.9$, resp., which are corresponding to standard deviation parameters of the Gaussian function in units of degrees of visual angle. The marginal posteriors of both parameters are largely overlapping. The exponent $\gamma \approx 1$ indicates that the weighting of the corresponding activation maps is negligible. Finally, the shift parameter η is clearly smaller than one, as expected.

There are two indicators lend support to the stability of our parameter estimations using the baseline SceneWalk model and the extended model in combination with the DREAM method. Firstly, the three chains which we ran for each subject resemble each other quite closely. Also, while the parameter estimates vary between participants, in most cases they do not differ dramatically (see Fig. A.7). Secondly, we conducted recovery analyses of the estimated parameters, where a parameter estimation is run on simulated data. The dream algorithm was able to identify the parameter values from the simulated data (see parameter recovery).

1.2.5 Individual parameter estimates

Each subject's data was individually fitted to both the baseline model and the extended model. In Table A.2 we report the results of parameter estimation of the baseline model (SceneWalk). Table A.3 gives the estimated parameters for the extended model. For each estimated parameter we computed the point estimate and the corresponding 50% credibility interval.

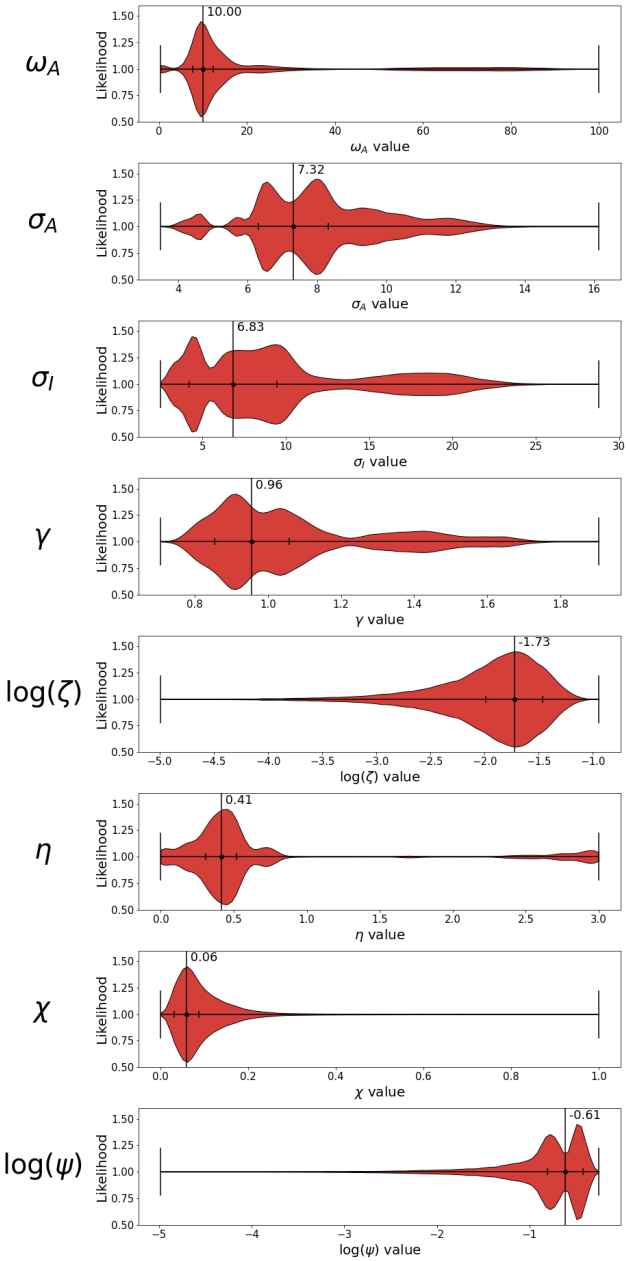

Figure A.6 Marginal posteriors of the estimated model parameters as calculated by fitting the model to the training data using PyDream.

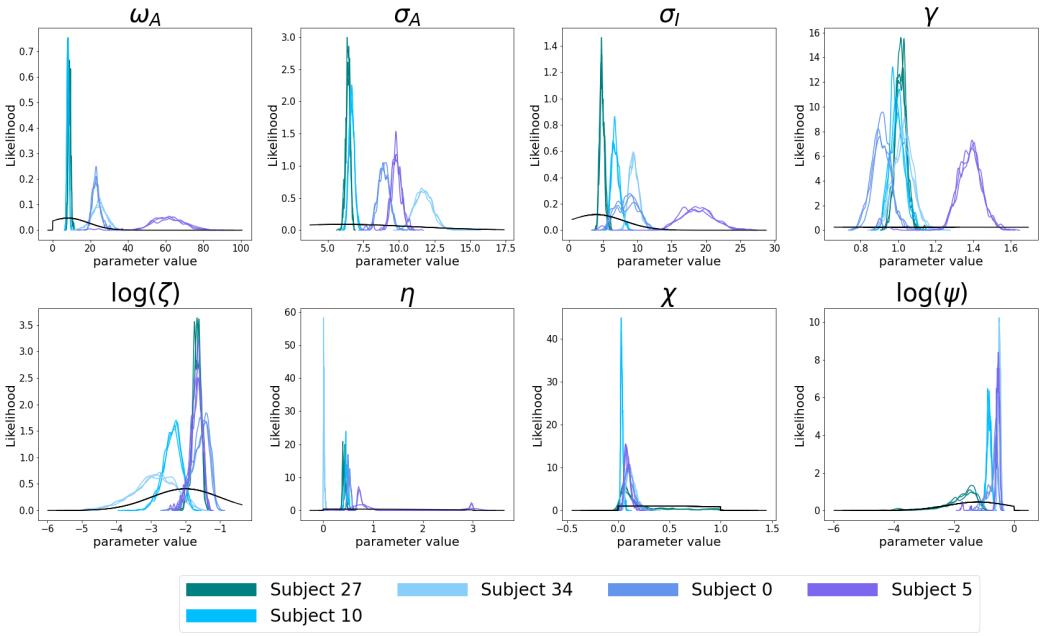

Figure A.7 Posterior density of individuals and chains of the model fit using DREAM. The different colors show chains belonging to one subject. The black line is the prior. The consistency of the chains within each subject indicates that we achieved a reliable fit. All chains are markedly different from the posterior, allowing us to update our beliefs.

Subject	γ	γ +/-	ω_A	ω_A +/-	σ_A	σ_A +/-	σ_F	σ_F +/-	$log(\zeta)$	$log(\zeta)$ +/-
0	0.927	0.025	24.368	1.702	7.551	0.141	5.566	0.264	-1.053	0.041
1	0.911	0.026	12.819	0.825	7.919	0.134	3.144	0.103	-1.172	0.057
2	1.082	0.023	17.484	1.154	6.406	0.101	5.111	0.208	-1.162	0.044
3	0.752	0.019	15.499	0.919	6.605	0.101	3.760	0.212	-1.402	0.072
4	1.409	0.030	44.364	2.301	9.069	0.183	0.061	0.014	-1.974	0.270
5	1.070	0.020	12.667	0.921	8.067	0.175	4.651	0.232	-1.343	0.058
6	0.963	0.020	14.522	0.791	5.137	0.085	3.333	0.181	-1.178	0.047
7	1.091	0.026	11.743	0.660	4.054	0.064	3.744	0.128	-1.056	0.031
8	0.741	0.014	18.658	1.272	5.595	0.090	5.228	0.178	-1.104	0.045
9	0.990	0.019	10.869	0.631	5.632	0.072	3.728	0.098	-1.697	0.077
10	0.960	0.027	12.875	0.925	6.161	0.097	5.844	0.224	-1.651	0.083
11	1.141	0.027	21.191	1.683	8.478	0.145	6.282	0.267	-1.091	0.041
12	0.935	0.021	14.712	1.112	4.923	0.090	3.760	0.228	-1.093	0.040
13	0.867	0.020	13.251	0.840	4.893	0.078	4.288	0.154	-1.366	0.050
14	0.810	0.018	15.372	0.929	7.214	0.109	4.275	0.190	-1.608	0.098
15	1.245	0.026	16.038	1.688	7.659	0.159	8.140	0.803	-1.106	0.035
16	0.982	0.029	21.626	2.352	4.775	0.077	5.497	0.234	-1.028	0.033
17	0.913	0.018	13.101	0.996	7.977	0.136	5.982	0.255	-1.896	0.136
18	0.978	0.020	16.447	1.258	8.533	0.166	6.587	0.333	-1.225	0.051
19	0.917	0.018	12.971	0.879	6.828	0.120	4.256	0.164	-1.199	0.050
20	0.958	0.033	13.985	1.574	8.016	0.177	7.987	0.502	-1.040	0.041
21	0.943	0.014	8.820	0.802	6.920	0.136	6.137	0.483	-1.496	0.062
22	1.367	0.029	15.801	0.892	8.700	0.135	3.196	0.110	-0.971	0.035
23	0.952	0.038	13.998	1.466	6.886	0.153	16.378	1.585	-1.045	0.044
24	1.167	0.021	15.325	1.163	6.360	0.104	5.206	0.277	-1.266	0.044
25	1.067	0.018	16.058	1.094	7.538	0.116	5.264	0.250	-1.384	0.054
26	1.229	0.030	8.680	0.869	9.245	0.269	5.290	0.271	-1.213	0.079
27	0.859	0.018	18.938	1.356	4.884	0.098	4.240	0.210	-1.525	0.085
28	1.196	0.019	20.110	1.568	8.180	0.142	5.410	0.199	-1.170	0.043
29	1.463	0.036	13.326	2.919	10.551	0.243	16.203	1.534	-1.076	0.036
30	0.920	0.019	15.208	0.947	7.034	0.122	3.822	0.132	-1.201	0.051
31	0.231	0.011	43.633	3.538	2.939	0.091	3.158	0.179	-0.977	0.045
32	0.862	0.023	16.398	1.213	3.920	0.070	2.775	0.122	-1.294	0.048
33	1.067	0.024	14.702	1.038	5.711	0.076	3.825	0.155	-1.261	0.045
34	0.944	0.032	24.298	2.569	9.347	0.227	6.307	0.349	-0.898	0.043

Table A.2 Estimated parameter values of the baseline SceneWalk model for all participants (see Table A.1 for fixed parameters).

Subject	χ	χ +/-	η	η +/-	γ	γ +/-	ω_A	ω_A +/-	σ_A	σ_A +/-	σ_F	σ_F +/-	$log(\psi)$	$log(\psi)$ +/-	$log(\zeta)$	$log(\zeta)$ +/-
0	0.090	0.027	0.511	0.022	0.905	0.032	22.820	1.305	8.941	0.268	9.185	1.177	-0.611	0.067	-1.460	0.158
1	0.066	0.030	0.441	0.015	0.904	0.026	8.661	0.714	9.180	0.187	3.426	0.234	-0.958	0.285	-1.493	0.139
2	0.082	0.024	0.413	0.018	0.911	0.025	15.505	0.981	7.104	0.181	7.282	0.460	-0.758	0.099	-1.689	0.144
3	0.055	0.020	0.493	0.019	0.802	0.022	12.810	0.642	7.532	0.168	4.458	0.213	-1.083	0.103	-1.673	0.092
4	0.175	0.043	2.974	0.026	1.339	0.048	79.172	3.848	8.326	0.219	18.587	1.275	-0.445	0.027	-2.102	0.214
5	0.081	0.019	0.727	0.040	1.385	0.040	59.578	5.479	9.725	0.226	18.229	1.609	-0.539	0.034	-1.646	0.104
6	0.064	0.015	0.417	0.009	1.033	0.026	11.031	0.488	6.813	0.151	6.111	0.490	-0.836	0.067	-1.632	0.094
7	0.118	0.021	0.494	0.018	1.444	0.042	8.440	0.462	4.217	0.159	15.433	1.436	-0.349	0.025	-1.616	0.118
8	0.056	0.008	0.428	0.021	1.051	0.026	9.341	0.514	7.569	0.198	9.563	0.636	-0.471	0.027	-1.864	0.140
9	0.041	0.041	0.509	0.016	1.099	0.018	8.598	0.334	6.379	0.113	3.151	0.111	-1.969	0.304	-1.730	0.077
10	0.031	0.007	0.463	0.018	0.992	0.025	7.980	0.361	6.655	0.127	6.689	0.376	-0.827	0.047	-2.362	0.167
11	0.102	0.033	0.399	0.014	0.932	0.021	12.103	0.949	9.466	0.258	8.558	0.391	-0.874	0.053	-1.508	0.072
12	0.060	0.010	0.307	0.017	1.084	0.026	10.367	0.703	6.418	0.146	12.359	0.723	-0.451	0.018	-2.280	0.219
13	0.065	0.014	0.522	0.013	0.899	0.020	9.628	0.399	5.692	0.098	6.871	0.310	-0.793	0.043	-1.949	0.110
14	0.085	0.048	0.583	0.022	0.810	0.023	14.693	0.740	7.931	0.120	4.586	0.162	-1.327	0.198	-1.947	0.145
15	0.094	0.019	0.318	0.025	0.908	0.022	7.679	0.666	8.342	0.217	10.099	0.410	-0.688	0.026	-2.845	0.379
16	0.069	0.012	0.350	0.009	1.284	0.028	10.213	0.503	6.709	0.172	17.440	1.361	-0.513	0.025	-1.411	0.058
17	0.038	0.034	0.382	0.025	0.886	0.020	13.462	1.033	8.223	0.162	6.178	0.298	-1.446	0.249	-1.715	0.091
18	0.044	0.009	0.323	0.011	1.044	0.022	9.492	0.546	11.713	0.375	9.008	0.789	-0.656	0.038	-1.908	0.127
19	0.114	0.089	0.302	0.019	0.866	0.018	8.752	0.679	7.804	0.172	4.349	0.146	-1.486	0.304	-1.385	0.066
20	0.077	0.022	0.751	0.019	0.982	0.037	29.371	3.378	8.213	0.200	9.908	0.831	-0.729	0.055	-1.264	0.064
21	0.060	0.014	0.368	0.022	1.024	0.028	10.314	0.958	8.103	0.164	9.421	0.562	-0.810	0.045	-1.901	0.113
22	0.147	0.023	2.816	0.136	1.469	0.036	70.456	3.942	10.353	0.248	15.990	0.947	-0.442	0.021	-2.528	0.347
23	0.051	0.007	0.198	0.019	1.402	0.049	0.625	0.040	10.185	0.679	18.559	1.445	-0.426	0.021	-1.946	0.129
24	0.108	0.024	0.164	0.012	1.160	0.023	9.564	0.448	8.012	0.199	7.788	0.357	-0.756	0.036	-1.984	0.131
25	0.030	0.005	0.037	0.023	0.827	0.020	12.721	1.070	7.789	0.203	8.460	0.536	-0.816	0.039	-2.485	0.222
26	0.052	0.014	0.161	0.023	1.133	0.032	6.289	0.558	11.923	0.545	9.724	0.713	-0.742	0.042	-2.161	0.225
27	0.072	0.046	0.392	0.021	1.016	0.019	8.630	0.448	6.435	0.098	4.820	0.217	-1.505	0.253	-1.667	0.080
28	0.067	0.010	2.529	0.121	1.621	0.049	84.455	4.655	9.431	0.238	19.496	1.182	-0.374	0.015	-2.264	0.205
29	0.157	0.029	2.755	0.062	1.621	0.042	60.145	4.287	10.576	0.265	20.243	1.272	-0.507	0.019	-1.955	0.139
30	0.035	0.010	0.483	0.032	0.828	0.018	10.687	0.555	7.496	0.163	4.320	0.265	-1.129	0.092	-1.500	0.077
31	0.071	0.013	0.687	0.011	0.934	0.024	13.099	0.618	6.988	0.141	9.374	0.539	-0.524	0.024	-2.094	0.167
32	0.031	0.009	0.488	0.017	0.919	0.018	9.758	0.409	4.633	0.076	3.845	0.159	-1.017	0.080	-1.718	0.092
33	0.057	0.014	0.253	0.011	0.941	0.025	18.227	1.651	6.418	0.144	6.284	0.580	-0.811	0.076	-1.785	0.110
34	0.111	0.027	0.013	0.006	1.030	0.035	24.958	2.591	11.618	0.429	9.267	0.477	-0.486	0.028	-2.767	0.402

Table A.3 Estimated parameter values of the extended SceneWalk model for all participants (see Table A.1 for fixed parameters).

B Appendix for Paper 2

B.1 Experimental details

2.1.1 Methods

The eye tracking setup included a mobile eye tracker in a lab with a wide projector screen. Subjects received credit points or a monetary compensation of 10,00€ for their participation. To increase compliance with the task, we offered participants an additional incentive of up to 3,00€ for correctly answering questions after each image (a total of 60 questions). The experiment was carried out in accordance with the Declaration of Helsinki. Informed consent was obtained for experimentation from all participants. The experiment data originally published by Backhaus et al. (2020) are freely available via OpenScienceFramework (OSF, https://osf.io/gxwfk/).

2.1.2 Data preprocessing

In our laboratory, we developed a processing workflow for the preprocessing of mobile eye-tracking data. Eye movement recordings from our mobile eye tracker are provided in head-centered coordinates. We presented 12 different QR codes around the stimulus material during the experiment. In the video output from the mobile eye tracker, we detected the QR codes using the Pupil Labs software Pupil Player version 1.7.42 (Kassner et al., 2014). The stimulus area within the QR codes is defined as a rectangle. Using a projective transformation provided by the image processing toolbox from MATLAB (The MathWorks, Natick/MA), we converted data points from head-centered coordinates (indicating points in the video frames) to image-centered coordinates (referring to the stimulus images).

After truncating the data to the relevant time segments of the stimulus presentation, we used a velocity-based saccade detection algorithm (Engbert & Mergenthaler, 2006; Engbert & Kliegl, 2003). For more detailed information on how to fit the parameters to our measurement device, please see Backhaus et al. (2020), where a number of filter criteria are described in detail. These criteria produce reliable data points, when working with the SMI Eye Tracking Glasses (SMI-ETG 2W; SensoMotoric Instruments, Teltow, Germany). After preprocessing, a total of 40,182 fixations and

47,425 saccades were retained for further analyses and modeling.

2.1.3 Most important results

The original experiment by Backhaus et al. (2020) reports statistics, from which we summarize the most relevant effects in the following. The authors looked at temporal and spatial eye movement parameters and compared the 4 different tasks using linear mixed models. The contrasts of the linear mixed models were chosen in such a way that the differences between the task groups (Guess conditions/free viewing vs. Count conditions/search) as well as the differences between the two specific tasks within a type could be compared (Guess time vs. Guess country; Count people vs. Count animals).

The authors report variations in fixation durations induced by the experimental task manipulations. On average, fixation durations are shorter in Count tasks compared to Guess tasks. Particularly short fixation durations occur in difficult Count tasks; Couting animals involves more challenging search components than counting people. Results also showed differences in saccade amplitudes between task types: Count tasks lead to shorter saccade amplitudes than Guess tasks. For saccade amplitudes, unlike fixation durations, no differences were found within task types. Backhaus et al. (2020) report that the tasks produced differences in gaze behavior on other spatial parameters. In Count tasks, participants disengaged faster and further from the image center (after generating the initial tendency to fixate the image center) compared to Guess conditions (Rothkegel et al., 2017; Tatler, 2007).

With respect to the image-dependent 2D density of fixations, gaze in the Count people condition focused on comparatively fewer salient locations while fixation locations in the Count animals condition were most distributed across the image. The Guess tasks induced distributions between these two extremes. Thus, there was a strong influence of the task on image-dependent saliency.

B.2 SceneWalk model specification

In the main text, we introduced the basic components of the SceneWalk model in its most recent version (Schwetlick, Rothkegel, Trukenbrod, et al., 2020b). We provide additional mathematical details in this appendix. As explained in the main text, the SceneWalk model comprises two largely independent processing streams, activation and inhibition, which when combined are interpreted as the fixation probability π at each grid point i, j at time t. In the original formulation of the model (Engbert, Trukenbrod, et al., 2015), the center of both the activation and the inhibition stream align with the current fixation position (f_x, f_y). The differential equations that define the temporal evolution of the activations of the two streams are given in Eq. (3.6) for the activation stream and in Eq. (3.5) for the inhibitory stream in the main text.

Over time intervals with constant input (i.e., during fixation, a closed-form solution

can be found by integrating analytically, i.e., for the activation

$$A(t) = \frac{G_A S}{\sum G_A S} + e^{-\omega_A (t-t_0)} \left(A_0 - \frac{G_A S}{\sum G_A S} \right), \tag{B.1}$$

and

$$F(t) = \frac{G_F}{\sum G_F} + e^{-\omega_F (t-t_0)} \left(F_0 - \frac{G_F}{\sum G_F} \right), \tag{B.2}$$

for the inhibition, where we dropped the indices i, j for simplification of the notation. It is important to note that the assumption of constant input is an approximation because of the presence of miniature eye movement produced involuntarily during fixation (e.g., Engbert & Mergenthaler, 2006).

The weighted difference of the activations in the two streams represents the priority map for target selection (Eq. (3.8). Since the difference will lead negative activations at locations, we take the part of the map, i.e.,

$$u^*_{ij}(u_{ij}) = \begin{cases} u_{ij}, & \text{if } u_{ij} > 0 \\ 0, & \text{otherwise.} \end{cases} . \tag{B.3}$$

The most recent version of the SceneWalk model (Schwetlick, Rothkegel, Trukenbrod, et al., 2020b) introduced different phases of perisaccadic influences during each fixation. Specifically, before and after a saccade, the center of the activation stream shifts. A pre-saccadic shift to the upcoming target occurs before saccade onset and post-saccadic shift in the direction of the saccade vector occurs after the saccade (Fig. 3.2). Thus, for a time τ_{pre} before each saccade, once the next location has been selected from the priority map with probability $\pi(i, j)$, the center of the Gaussian input shifts to the location of the upcoming fixation, i.e.,

$$G_A^{pre}(x, y) = \frac{1}{2\pi\sigma_A^2} \exp\left(-\frac{(x - x_{f+1})^2 + (y - y_{f+1})^2}{2\sigma_A^2} \right) , \tag{B.4}$$

When the pre-saccadic phase terminates, the saccade is executed. For the purposes of this work, we neglect saccade durations, as most information is acquired during fixations. Now, the post-saccadic shift phase begins, during which the center of the activation Gaussian is determined by Eq. (3.7). The evolution equation is then given by

$$G_A^{post}(x, y) = \frac{1}{2\pi\sigma_{post}^2} \exp\left(-\frac{(x - x_s)^2 + (y - y_s)^2}{2\sigma_{post}^2} \right) . \tag{B.5}$$

As the inhibition stream always aligns with the fixation location, it can still be calculated for the entire fixation duration via Eq. (B.2). The result of the phase-specific activation and inhibition can be combined at any point in time to yield the fixation selection probability at that time.

Facilitation of return is implemented in the model as a selectively slower decay of attention ω_A at the one back location. It thus occurs more briefly and at a different

time scale than the inhibition of return implemented in the inhibition stream. The reduced decay rate $\omega_{\mathrm{F}oR}$ occurs in a spatial window $x - \nu < x_{f-1} < x + \nu$ and $y - \nu < y_{f-1} < y + \nu$ around the previous fixation location (x_{f-1}, y_{f-1}), where ν is the size of the window. We then replace ω_A in the evolution equation with a matrix that contains the value of ω_A everywhere except in the specified window, where it contains $\omega_{\mathrm{F}oR}$

$$A(t) = \frac{G_A S}{\sum G_A S} + e^{-\omega_{FoR}(t-t_0)} \left(A_0 - \frac{G_A S}{\sum G_A S} \right). \tag{B.6}$$

As suggested by Rothkegel et al. (2017), starting the model with a central activation improves the predictions of the model. Initially we instantiated the model with uniform distributions. The implementation of a transient central fixation bias changes the evolution equation for the first fixation so that

$$A(t) = \frac{G_{fix} S}{\sum G_{fix} S} + e^{-\omega_{cb}(t-t_0)} \left(A_{0_{CB}} - \frac{G_{fix} S}{\sum G_{fix} S} \right). \tag{B.7}$$

Finally, we implemented an additional bias towards horizontal and verical saccade directions (Engbert et al., 2011). The oculomotor map is centered at the current fixation location, i.e.,

$$P_{OM} = \left((x - x_f)^2 \cdot (y - y_f)^2 \right)^{\chi} , \tag{B.8}$$

where the factor χ determines the steepness of the oculomotor potential. In this variation, before the normalization and the addition of noise, Eq. (B.3, 3.10), the oculomotor map is added as

$$u_{\mathrm{O}M} = u + \left(\psi \cdot \left| \frac{P_{OM}}{\max(P_{OM})} - 1 \right| \right) , \tag{B.9}$$

where $\psi = 10^{-0.6}$ is a constant parameter.

B.3 Bayesian inference workflow

In this paper we applied a Bayesian inference workflow to a biologically plausible generative model. This approach is extremely promising for cognitive modeling for four reasons illustrated in the infographic in Fig. B.1.

In this framework a model is defined by its likelihood function and parameters. It can be used to calculate the probability of a given data point. Given a starting point it can also be used generatively to simulate data. Both the predictive and the generative parts of the model are necessary components of the proposed workflow and provide valuable insight into the model's characteristics.

First, we use the model likelihood to estimate the best values for the model parameters using Bayesian inference. The Bayesian parameter estimation algorithm repeatedly computes the model likelihood given the data, while systematically varying the parameter values. Thus, it tries to maximize the performance of the model using

the likelihood given the data. This process yields marginal posterior distributions for each parameter. These marginal posteriors can be interpreted as a rich source of information about the parameter as shown in box (c).

Second, we parametrize the model with the values obtained from the estimation and use it to simulate data. When fitting a model using an ad-hoc loss function, the model is trained specifically to reproduce whatever the chosen metrics may be. By contrast, using the likelihood allows for greater generalizability as well as avoiding overfitting. Simulated data can be compared to experimental data in order to assess how well the model reproduced trends that it was not directly informed of. To this end we perform a series of posterior predictive checks, which ascertain whether the model can actually capture the relevant features found in the data. Thus, they reveal strengths and weaknesses of the model regarding its plausibility.

Lastly, the model likelihood is relevant also for inter-model comparisons. It is a fair basis for comparison, in the sense that it provides the same information to each model with the experimental data. Each model can be fitted and compared in the same way: estimation algorithms determine the parameters using a training set of experimental data. Then, using a test set of experimental data, we can calculate and compare their performance.

B.4 Convergence of parameter estimation

As suggested by the authors of the DREAM algorithm (Vrugt & Braak, 2011), we used the Gelman-Rubin convergence diagnostic $\hat{R}$ to determine adequate quiality of the parameter estimation. The results are illustrated below for for all 256 models (Figure B.2). We used the value of 1.05 as a threshold to indicate convergence. In total of the 2304 fitted parameters, 2288 converged and 16 did not converge. At the level of models, of the 256 fitted models there were only 3 where the posterior did not converge for one or more parameters.

B.5 Preregistration

This work was preregistered at the Open Science Framework (OSF)[7] (Schwetlick, Backhaus, & Engbert, 2020) using the "Preregistration Template for the Application of Cognitive Models" (Crüwell & Evans, 2019). Please refer to the OSF repository for full information on the preregistration. Here we would like to follow up on some aspects of the preregistration and explain where and why we deviated from the preregistered research plan.

The hypotheses we stated in the preregistration concerned (a) differences in model parameters relating to the attention span for the different tasks and (b) the importance of inhibition for different tasks. For the former hypothesis our findings agreed that the

7 See https://osf.io/79qy8

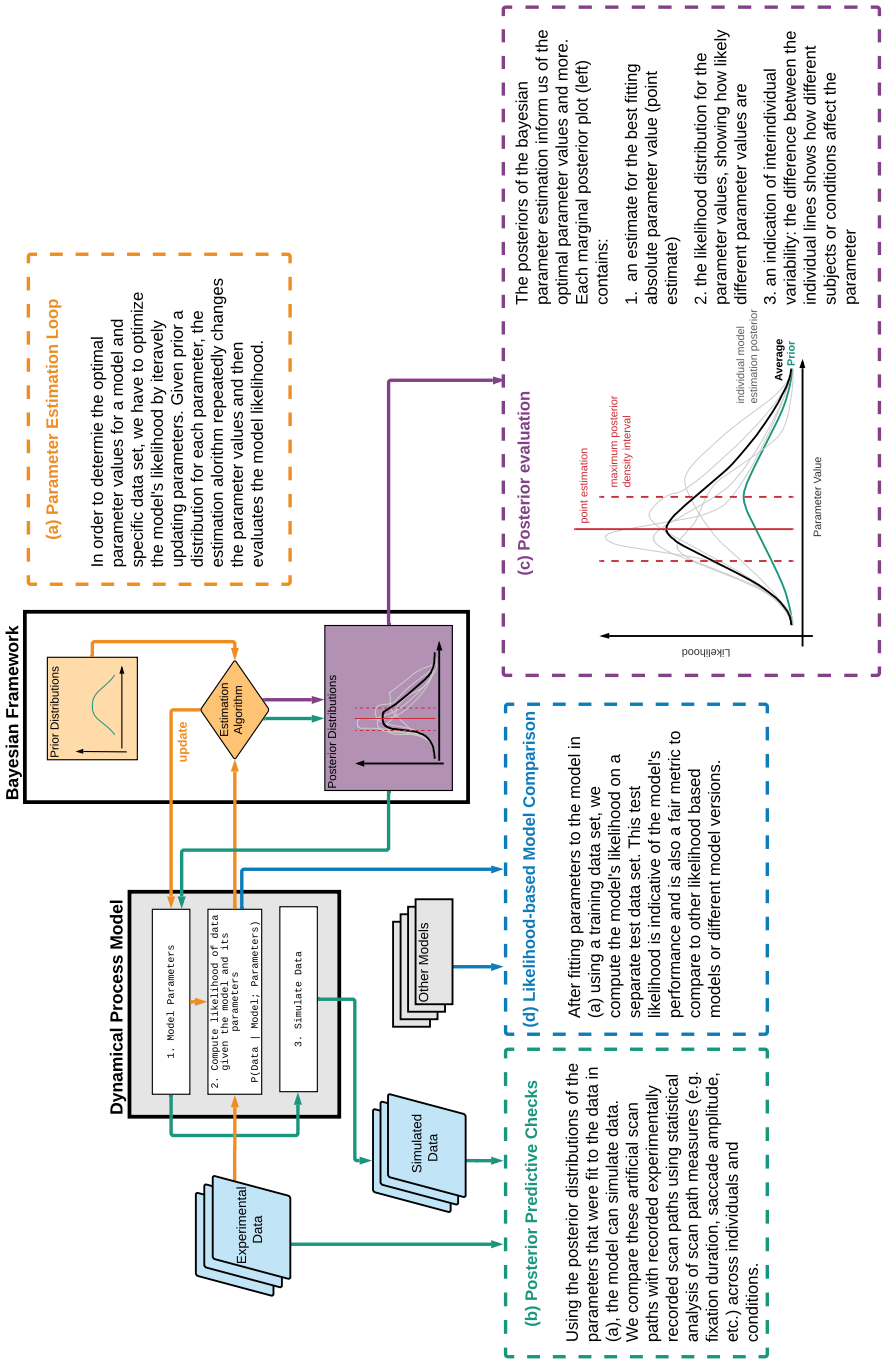

Figure B.1 Workflow for likelihood-based Bayesian inference. The workflow summarizes all steps of Bayesian inference and highlights four core advantages of the proposed workflow. Orange arrows and lines (a) refer to the statistically rigorous estimation of parameters using the model's likelihood function and empirical data. Green arrows and lines (b) show the process of conducting posterior predictive checks, where the resultant models' predictions are evaluated against real world data. Purple arrows and lines (c) explain how the specific parameter posteriors can be interpreted in a biologically-founded model. Lastly, blue arrows and lines (d) explain how the method is useful to establish comparability between competing models.

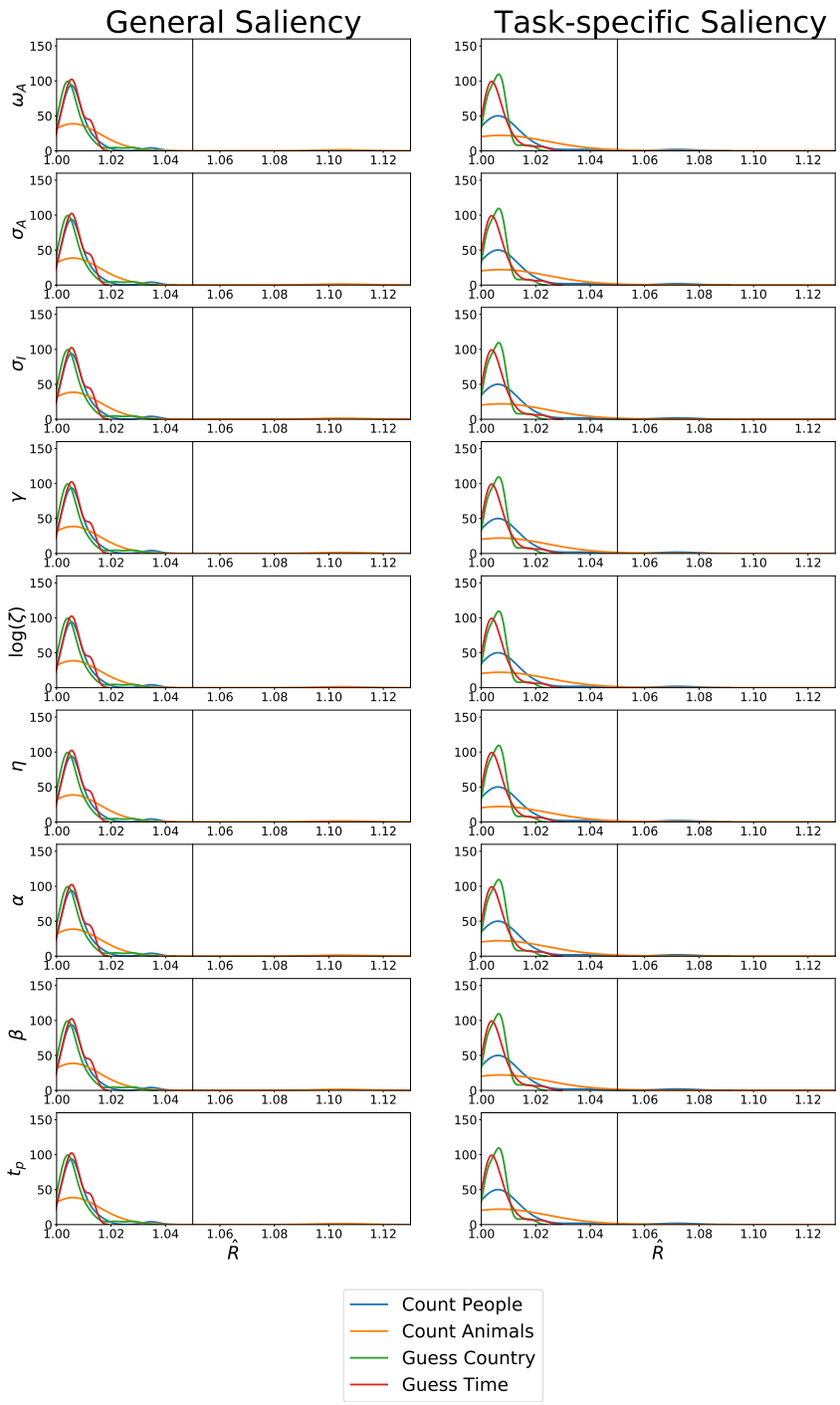

Figure B.2 Rhat Convergence. The plot shows the distribution of $\hat{R}$ values over our models. Each panel row represent a parameter. The horizontal line indicates the value of 1.05, a common threshold for determining convergence.

attentional span is greater in Guess task conditions than in Count tasks. As predicted, the activation Gaussian σ_A is greater in free viewing-like tasks.

We also find support for the latter hypothesis: In the Count task conditions, the span for the fixation map σ_F is smaller than in the Guess tasks, showing at least a more focused, localized inhibition component. The parameter C_F, which is mentioned as an exploratory analysis target in the preregistration was not included as a free parameter in the final estimation, as it turned out to be more difficult to identify given the relatively small amount of data available for each model fit.

The third hypothesis in the preregistration concerns parameter ω_A, which controls the speed of decay. We predicted a smaller value of ω_A for Count tasks, as we thought that keeping track of past fixations would be of greater use. We found this to be true, but only for models fitted using the general saliency map, not for models using task-specific saliency. We propose potential reasons for this finding in the discussion.

We also proposed a Markov-order analysis of the model to determine the influence of past states on the current. This analysis is not included in the current manuscript, since pilot simulations indicated that the analysis required larger amounts of data per fit than available from the current study. However, we consider the mathematical concept promising and aim to include a corresponding analysis in future work. The same is true for the mean-lag distance analysis proposed in the preregistration.

An important point in the preregistration was the possibility of running model fits based on individual data sets per task. As this was successful despite the limited data, the results of the current work are exclusively based on this strategy of fitting data for individuals and tasks independently. The alternative proposal of fitting models for each task by pooling across participants was no longer necessary. Additionally, instead of the proposed 5 free parameters for model fitting, we now successfully estimated 9 free parameters per data set, with 3 free parameters added the model to include the new temporal control of fixation durations.

B.6 Additional results

The following section provides additional details concerning the statistical analyses presented in this paper. Specifically, we provide the detailed results of our LMM analyses. Table B5 summarizes all of the applied LMM model structures.

The first LMM analysis compares the likelihoods of different versions of the SceneWalk model and Density Sampling models. The results can be found in Table B.1. The second analysis comprises an LMM of the posterior of each of the 9 estimated parameters for both general and task-specific model variants, i.e., 18 separate models. Fixed and random effects for each LMM are reported in Table B.2. Finally, Tables B3 and B4 contain the results of the LMMs pertaining to the comparison of simulated and experimental saccade amplitudes and fixation durations, resp.

	Custom contrast				Treatment contrast		
	β	SE	t		β	SE	t
Fixed Effects							
Intercept	15.10	0.066	228.74	T1	15.07	0.069	218.72
FModel	0.34	0.023	15.01	T2 - T1	0.31	0.032	9.72
FSal	0.25	0.023	11.04	T3 - T1	-0.28	0.032	-8.70
FInter	-0.06	0.046	-1.26	T4 - T1	0.09	0.032	2.80
	Var	SD			Var	SD	
Random Effects							
Subject: Intercept	0.0327	0.181			0.0327	0.181	
Image: Intercept	0.0960	0.310			0.0960	0.310	
Residual	0.2205	0.470			0.2205	0.470	
Number of obs	1700				1700		
Number of groups	subject	32			subject	32	
	image	30			image	30	

Table B.1 LMM fit by maximum likelihood - Comparison of the model likelihood gain. $|t| > 2$ are interpreted as significant effects, FModel: factor model 'Density Sampling' vs. 'SceneWalk', FSal: factor saliency 'General Saliency' vs. 'Task-specific Saliency', FInter: Factor for the interaction of FModel and FSal, T1: 'Task-specific Saliency - Density Sampling', T2: 'Task-specific Saliency - SceneWalk', T3: 'General Saliency - Density Sampling', T4: 'General Saliency - Scene Walk', To avoid zeros in the model likelihood gain, data were linearly transformed by adding 14.

	ω_A – Task-specific Saliency			ω_A – General Saliency		
Fixed Effects	β	SE	t	β	SE	t
Intercept	0.760	0.004	204.85	1.053	0.001	1102.15
FGC	-0.001	0.006	-0.22	0.006	0.002	2.60
FC	-0.014	0.011	-1.20	0.000	0.002	-0.07
FG	0.013	0.007	2.03	-0.006	0.002	-2.59
Random Effects	Var	SD		Var	SD	
Subject: Intercept	0.0004	0.021		0.0000	0.005	
Subject: FGC	0.0011	0.033		0.0001	0.012	
Subject: FC	0.0041	0.064		0.0002	0.014	
Subject: FG	0.0014	0.037		0.0002	0.014	
Residual	0.0005	0.022		0.0000	0.006	

	σ_A – Task-specific Saliency			σ_A – General Saliency		
Fixed Effects	β	SE	t	β	SE	t
Intercept	2.923	0.037	79.17	5.381	0.119	45.11
FGC	0.233	0.062	3.79	0.760	0.209	3.63
FC	0.139	0.076	1.82	0.682	0.226	3.01
FG	0.022	0.058	0.38	0.154	0.234	0.66
Random Effects	Var	SD		Var	SD	
Subject: Intercept	0.0436	0.209		0.4551	0.675	
Subject: FGC	0.1214	0.348		1.4028	1.184	
Subject: FC	0.1852	0.430		1.6362	1.279	
Subject: FG	0.1085	0.329		1.7575	1.326	
Residual	0.0352	0.188		0.3657	0.605	

	σ_F – Task-specific Saliency			σ_F – General Saliency		
Fixed Effects	β	SE	t	β	SE	t
Intercept	4.035	0.069	58.40	4.710	0.103	45.65
FGC	0.626	0.132	4.75	0.778	0.164	4.73
FC	0.449	0.169	2.65	0.836	0.211	3.97
FG	0.217	0.181	1.20	0.448	0.245	1.83
Random Effects	Var	SD		Var	SD	
Subject: Intercept	0.1526	0.391		0.3401	0.583	
Subject: FGC	0.5550	0.745		0.8629	0.929	
Subject: FC	0.9145	0.956		1.4193	1.191	
Subject: FG	1.0450	1.022		1.9103	1.382	
Residual	0.3575	0.598		0.6995	0.836	

Table B.2 LMM fit by maximum likelihood - Model parameter with our custom contrasts. FGC: first contrast 'Count' vs. 'Guess', FC: second contrast 'Count People' vs. 'Count Animals', FG: third contrast 'Guess Country' vs. 'Guess Time', $|t| > 2$ are interpreted as significant effects.

	γ_i – Task-specific Saliency			γ_i – General Saliency		
Fixed Effects	β	SE	t	β	SE	t
Intercept	1.000	0.000	3338.22	0.996	0.005	183.68
FGC	0.001	0.001	2.38	-0.005	0.010	-0.49
FC	0.000	0.001	-0.41	0.035	0.009	4.02
FG	0.000	0.001	-0.57	-0.003	0.016	-0.17
Random Effects	Var	SD		Var	SD	
Subject: Intercept	0.0000	0.002		0.0009	0.031	
Subject: FGC	0.0000	0.003		0.0031	0.056	
Subject: FC	0.0000	0.005		0.0024	0.049	
Subject: FG	0.0000	0.004		0.0085	0.092	
Residual	0.0000	0.002		0.0016	0.039	

	$\log_{10}\zeta$ – Task-specific Saliency			$\log_{10}\zeta$ – General Saliency		
Fixed Effects	β	SE	t	β	SE	t
Intercept	31.770	0.455	69.77	61.388	0.856	71.72
FGC	-4.502	0.980	-4.60	-11.040	2.398	-4.60
FC	-2.098	1.527	-1.37	-3.719	3.014	-1.23
FG	-0.858	0.960	-0.89	-0.482	2.758	-0.17
Random Effects	Var	SD		Var	SD	
Subject: Intercept	6.6174	2.572		23.3713	4.834	
Subject: FGC	30.6415	5.535		183.7608	13.556	
Subject: FC	74.5076	8.632		290.1551	17.034	
Subject: FG	29.3677	5.419		242.8092	15.582	
Residual	32.3881	5.691		125.8949	11.220	

	η – Task-specific Saliency			η – General Saliency		
Fixed Effects	β	SE	t	β	SE	t
Intercept	0.861	0.006	155.14	0.808	0.007	112.35
FGC	-0.005	0.013	-0.42	-0.004	0.016	-0.22
FC	-0.017	0.013	-1.29	-0.032	0.017	-1.92
FG	0.002	0.011	0.19	-0.001	0.016	-0.04
Random Effects	Var	SD		Var	SD	
Subject: Intercept	0.0010	0.031		0.0017	0.041	
Subject: FGC	0.0052	0.072		0.0085	0.092	
Subject: FC	0.0055	0.074		0.0090	0.095	
Subject: FG	0.0038	0.062		0.0078	0.088	
Residual	0.0013	0.035		0.0018	0.042	

Table B.2 (cont'd) LMM fit by maximum likelihood - Model parameter with our custom contrasts. FGC: first contrast 'Count' vs. 'Guess', FC: second contrast 'Count People' vs. 'Count Animals', FG: third contrast 'Guess Country' vs. 'Guess Time', $|t| > 2$ are interpreted as significant effects.

	t_α – Task-specific Saliency			t_α – General Saliency		
Fixed Effects	β	SE	t	β	SE	t
Intercept	1.749	0.022	80.28	2.108	0.026	81.29
FGC	0.272	0.044	6.18	0.024	0.060	0.39
FC	-0.067	0.060	-1.12	-0.113	0.080	-1.41
FG	0.022	0.052	0.42	0.127	0.069	1.85
Random Effects	Var	SD		Var	SD	
Subject: Intercept	0.0149	0.122		0.0210	0.145	
Subject: FGC	0.0606	0.246		0.1121	0.335	
Subject: FC	0.1132	0.336		0.2013	0.449	
Subject: FG	0.0836	0.289		0.1473	0.384	
Residual	0.5813	0.762		0.9381	0.969	

	t_β – Task-specific Saliency			t_β – General Saliency		
Fixed Effects	β	SE	t	β	SE	t
Intercept	73.913	0.569	129.87	75.114	0.575	130.71
FGC	0.996	0.904	1.10	-1.094	0.889	-1.23
FC	3.713	0.879	4.22	3.835	0.895	4.29
FG	-0.881	1.065	-0.83	-0.558	1.086	-0.51
Random Effects	Var	SD		Var	SD	
Subject: Intercept	10.3581	3.218		10.5603	3.250	
Subject: FGC	26.1320	5.112		25.2455	5.024	
Subject: FC	24.6481	4.965		25.5472	5.054	
Subject: FG	36.2232	6.019		37.7145	6.141	
Residual	13.7678	3.711		13.3568	3.655	

	q – Task-specific Saliency			q – General Saliency		
Fixed Effects	β	SE	t	β	SE	t
Intercept	2.598	0.028	93.99	2.574	0.028	91.44
FGC	0.017	0.038	0.45	0.074	0.039	1.89
FC	-0.178	0.041	-4.39	-0.181	0.042	-4.36
FG	0.053	0.054	0.97	0.046	0.056	0.82
Random Effects	Var	SD		Var	SD	
Subject: Intercept	0.0244	0.156		0.0253	0.159	
Subject: FGC	0.0466	0.216		0.0484	0.220	
Subject: FC	0.0525	0.229		0.0552	0.235	
Subject: FG	0.0948	0.308		0.0992	0.315	
Residual	0.0262	0.162		0.0261	0.161	

Table B.2 (cont'd) LMM fit by maximum likelihood - Model parameter with our custom contrasts. FGC: first contrast 'Count' vs. 'Guess', FC: second contrast 'Count People' vs. 'Count Animals', FG: third contrast 'Guess Country' vs. 'Guess Time', $|t| > 2$ are interpreted as significant effects.

	Experimental data			Simulated data General Saliency			Simulated data Task-specific Saliency		
	β	SE	t	β	SE	t	β	SE	t
Fixed Effects									
Guess - Count	0.10	0.009	10.66	0.05	0.009	4.89	0.07	0.009	7.23
CountAnimals - CountPeople	0.07	0.013	5.27	0.08	0.013	5.99	0.05	0.013	3.84
GuessTime - GuessCountry	0.00	0.014	-0.05	0.00	0.014	0.36	0.01	0.014	1.09
	Var	SD		Var	SD		Var	SD	
Random Effects									
Subject: Intercept	0.0072	0.085		0.0036	0.060		0.0053	0.073	
Image: Intercept	0.0167	0.129		0.0016	0.040		0.0015	0.039	
Residual	0.7338	0.857		0.7348	0.857		0.7225	0.850	
Number of obs	34188			33941			34089		
Number of groups	subject	32		subject	32		subject	32	
	image	30		image	30		image	30	

Table B3 LMM fit by maximum likelihood – Saccade amplitudes (log-transformed) for our contrasts. $|t| > 2$ are interpreted as significant effects.

	Experimental data			Simulated data General Saliency			Simulated data Task-specific Saliency		
	β	SE	t	β	SE	t	β	SE	t
Fixed Effects									
Guess - Count	0.02	0.005	4.73	0.03	0.005	5.05	0.04	0.005	8.37
CountAnimals - CountPeople	-0.05	0.007	-6.88	-0.04	0.007	-5.77	-0.02	0.007	-3.17
GuessTime - GuessCountry	0.02	0.007	3.39	0.01	0.008	0.75	0.03	0.008	3.87
	Var	SD		Var	SD		Var	SD	
Random Effects									
Subject: Intercept	0.0072	0.085		0.0081	0.090		0.0077	0.088	
Image: Intercept	0.0022	0.047		0.0003	0.018		0.0005	0.022	
Residual	0.2122	0.461		0.2364	0.486		0.2344	0.484	
Number of obs	34873			34873			34873		
Number of groups	subject	32		subject	32		subject	32	
	image	30		image	30		image	30	

Table B4 LMM fit by maximum likelihood – Fixation durations (log-transformed) for our contrasts. $|t| > 2$ are interpreted as significant effects.

Dependent variable		Fixed effect part	Random effect part
*Structure of LMM with custom contrast – Comparison of the model likelihood gain in Figure 3.4b**			
Model likelihood gain	$\sim$	$1 + \text{FModel} + \text{FSal} + \text{FInter}$	$+ \quad (1 \mid \text{subject}) \quad + \quad (1 \mid \text{image})$
*Structure of LMM with treatment constrast – Comparison of the model likelihood gain in Figure 3.4b**			
Model likelihood gain	$\sim$	$\text{T1} + \text{T2} + \text{T3} + \text{T4}$	$+ \quad (1 \mid \text{subject}) \quad + \quad (1 \mid \text{image})$
Structure of the 18 LMMs in Figure 3.7			
Model Parameter	$\sim$	$1 + \text{FGC} + \text{FC} + \text{FG}$	$+ \quad (1 + \text{FGC} + \text{FC} + \text{FG} \mid\mid \text{subject})$
*Structure of the three LMMs of log saccade amplitudes in Figure 3.8a**			
Log Saccade Amplitudes	$\sim$	$1 + \text{FGC} + \text{FC} + \text{FG}$	$+ \quad (1 \mid \text{subject}) \quad + \quad (1 \mid \text{image})$
*Structure of the three LMMs of log fixation duration in Figure 3.8b**			
Log Fixation Duration	$\sim$	$1 + \text{FGC} + \text{FC} + \text{FG}$	$+ \quad (1 \mid \text{subject}) \quad + \quad (1 \mid \text{image})$

Table B5 LMM model structure. 1: Intercept, FModel: factor model 'Density Sampling' vs. 'SceneWalk', FSal: factor saliency 'General Saliency' vs. 'Task-specific Saliency', FInter: Factor for the interaction of FModel and FSal, T1: 'Task-specific Saliency - Density Sampling', T2: T1 vs. 'Task-specific Saliency - SceneWalk', T3: T1 vs. 'General Saliency - Density Sampling', T4: T1 vs. 'General Saliency - Scene Walk', FGC: first contrast 'Count' vs. 'Guess', FC: second contrast 'Count People' vs. 'Count Animals', FG: third contrast 'Guess Country' vs. 'Guess Time', $||$: double bar sign represents that the correlations of random effects are not included in the model, *we choose the minimal model with only random intercepts for subjects and images to have comparable models between all subsets of this analysis.

C Preregistration for Paper 2

C.1 Study information

3.1.1 Title

Modeling task-dependence in natural scene viewing using the SceneWalk model of scan path generation

3.1.2 Authors

Lisa Schwetlick, Daniel Backhaus, Ralf Engbert

3.1.3 Description

The fixation locations chosen during visual perception of natural scenes depend on the observer's objective while viewing the image. Static models of eye movement behavior, can predict near-perfect predictions of fixation densities, which reflect the different tasks. The dynamic, temporal component is less well explored, however. Using a biologically plausible model of scan paths based on neurophysiological assumptions we want to model experimental data for different viewing tasks. We will investigate whether we can explain the difference between the tasks by varying model parameters that directly correspond to the biologically interpretable concepts in the visual system.

3.1.4 Hypotheses

Main Hypothesis: The SceneWalk model can represent differences between task conditions in scene viewing data 1. in terms of estimated parameter values, and 2. in terms of differences between generated scan paths

H1) We expect to find a difference between task conditions (Count People, Count Animals, Guess Time and Guess Country) concerning the attentional span. In free-viewing-like tasks (Guess Time and Guess Country) participants have an increased attentional span compared to search-like tasks (Count People, Count Animals). Previous research has shown that saccade amplitude and attentional span are related

and saccade amplitudes tend to be smaller in search tasks (Trukenbrod et al., 2019). Attentional span is most closely related to parameter σ_A in the model. (directional)

H2) In search-like tasks inhibitory tracking is more important than in free-viewing-like tasks. We expect parameters that strengthen the inhibition component to be more pronounced for search-like tasks than for free-viewing-like tasks. Stronger inhibition drives scene exploration to scan further areas in the scene. The inhibition path is defined mainly by parameters σ_I and CF in the model. (directional)

H3) In search-like tasks fixation history is more relevant, as it is necessary to keep track of previously visited locations. The model's memory span is determined by parameter ω. In search-like tasks we therefore expect a smaller value for the parameter ω. Additionally, the number of past fixations that add a benefit for predicting the next one can be investigated. We expect the limit to useful past fixations in search-like tasks to be greater than in free-viewing-like tasks.

C.2 Data description for pre-existing data

3.2.1 Name or brief description of dataset(s)

Name: MBody1 and MBody2

Eye tracking data from a scene viewing experiment using a mobile eye tracker in a lab with a wide projector screen. In Mbody1, the data set we will model, 32 participants viewed 30 images twice under 4 different concrete viewing task conditions. The tasks were:

1. Count People (search-like)
2. Count Animals (search-like)
3. Guess Time (free-viewing-like)
4. Guess Country (free-viewing-like).

In the following questions, we refer to this data set. In order to generate the fixation density maps needed for the model, we used the fixation data on each image and each task. Additionally, we generated free viewing fixation densities from a second data set on the same images: a subset of MBody2, where 32 participants viewed the same 30 images under a less concrete, free viewing instruction.

3.2.2 Is this data open or publicly available?

Yes, for MBody1

3.2.3 How can the data be accessed?

- **MBody1:** https://doi.org/10.17605/OSF.IO/GXWFK, https://osf.io/gxwfk/
- **MBody2:** data are not openly available yet.

3.2.4 Date of download, access, or future access

Not applicable

3.2.5 Data source

Own lab collection - The data were collected in 2017-2018 by Daniel Backhaus in the Eye Lab of the University of Potsdam, Germany.

3.2.6 Additional information about data source

Not applicable

3.2.7 Codebook

Not applicable

3.2.8 Sampling and data collection procedures

For this study, we used data of 32 students of the University of Potsdam with normal or corrected to normal vision. On average participants were 22.8 years old (18–36 years) and 31 participants were female. Participants received credit points or a monetary compensation of 10,00 €. To increase compliance with the task, we offered participants an additional incentive of up to 3,00€ for correctly answering questions after each image (in sum 60 questions). The work was carried out in accordance with the Declaration of Helsinki. Informed consent was obtained for experimentation from all participants.

For further information see https://doi.org/10.1167/jov.0.0.06824, https://arxiv.org/abs/1911.06085.

3.2.9 Prior work based on the dataset

- Paper published in Journal of Vision, preprint available on https://arxiv.org/abs/1911.06085, https://doi.org/10.1167/jov.0.0.06824
- Conference Poster Presentation: 20th European Conference on Eye Movements ECEM 2019
- Scandinavian Workshop on Applied Eye Tracking - SWAET. 2018

3.2.10 Prior research activity

All the research we have done on this data is included in point C.2

3.2.11 Prior knowledge of the current dataset

DB and RE have experimentally analyzed the data. The data has not previously been used for modeling and LS is unfamiliar with the data. For experimental results please see https://arxiv.org/abs/1911.06085, https://doi.org/10.1167/jov.0.0.06824

The core findings concerning task differences relevant to this project are summarized here:

- Saccade amplitudes: Free-viewing-like tasks produce longer saccade amplitudes than search like tasks
- Free-viewing-like tasks produce a greater shannon's entropy.
- Count Animals produces a greater shannon's entropy than Count People tasks.
- Guess Time task produces a greater shannon's entropy than Guess Country task
- Count Animals task produces a smaller predictability than Count People task.
- Guess Time task produces a smaller predictability than Guess Country task

C.3 Sampling plan

Not applicable; as in existing data.

C.4 Design plan

Not applicable; as in existing data.

C.5 Variables

Not applicable; as in existing data.

C.6 Data cleaning and preparation

3.6.1 Data exclusion

See https://arxiv.org/abs/1911.06085, https://doi.org/10.1167/jov.0.0.06824

Further, in order to generate the empirical fixation density maps we only used a subset of MBody2 where participants had a free viewing task. Besides that the same filter criteria was used as described in the paper. The eye movement data contain measurement error and noise, most prominently through eye blinks. Blink data points were removed from the data used in the model.

3.6.2 Data partitioning for train/test

We will split the data into 4 sets, one for each of the 4 viewing task conditions. Each of the 4 sets will be split into test and training data randomly such that 75% of the data of each participant is in the training data set and 25% is in the test data set.

In an exploratory analysis we will compute individual fits for each subject, in which case the data will be split by individual and by each of the 4 viewing task conditions. This will result in 32*4=128 data sets. The test and training split will be as described above.

C.7 Modeling

3.7.1 Mathematical or computational model

We will be modeling the data using the extended SceneWalk model (https://psyarxiv.com/zcbny/) Short Summary of the Model: The SceneWalk model is a dynamical model of fixation selection that relies on two separate processing streams. The attention stream combines visual saliency with a gaussian blob centered around the current fixation location in order to simulate the decrease in visual accuracy in the peripheral visual field. The inhibition stream drives eye movement away from the current fixation location by another gaussian blob around the current fixation location. Both streams are implemented on a 128x128 grid, evolve independently over time, and are finally subtracted in order to yield a priority map from which the next fixation is sampled. In a recent study we extended the SceneWalk model to include mechanisms of perisaccadic attention shifts, which allow the model to better capture systematic tendencies found ubiquitously in scan path data. Specifically the extended model implements a mechanism of facilitation of return, pre- and post-saccadic attentional shifts and oculomotor potential.

Parameters: For this study we will estimate the following relevant parameters:

- σ_A
- σ_I
- ω
- ζ
- η

All other model parameters will be set to default values determined by previous work.

Priors: We use Bayesian parameter inference to fit the model. Priors are informed by the previous study (https://psyarxiv.com/zcbny/).

Data: The model fit is computed on a set of training data. This model predicts scan path dynamics and requires baseline saliency information to be provided. For this

purpose we use empirical fixation density maps. In a first step we will train the model using fixation density maps that are specific to the image and the task (generated from MBody1 data). In a second step we will explore if scan path dynamics are sufficient to explain the difference in behavior between tasks by using general fixation density maps based on free viewing behavior from another study (MBody2).

Implementation: The model implementation can be found at https://github.com/lschwetlick/SceneWalk_Model

3.7.2 Method of parameter and hyperparameter estimation

Estimation Procedure: The SceneWalk model is a likelihood-based model which allows us to estimate parameters directly without relying on ad-hoc performance metrics. The model definition and selection of parameters to be estimated is informed by the previous study using the same model. The parameters we do not estimate were fixed either because they could not be estimated due to their small effect or because they were found, in separate analyses using different data, to destabilize the model. The parameters are estimated using the Differential Adaptive Metropolis Sampler in Python (PyDream; https://github.com/LoLab-VU/PyDREAM).

Separate Fits: We will fit parameters to data from each of the four tasks. In an exploratory analysis, we will fit each subject and task individually. The uncertainty in this analysis is whether the dataset will be too small to allow parameters to converge.

C.8 Robustness checks and model testing

3.8.1 Robustness checks and sensitivity analyses

We will verify the robustness of our fits using A qualitative check of convergence. The posterior chains plots of the DREAM sampler will provide information on whether the estimated parameters could be constrained by the data. Performing a recovery analysis of a data set. This analysis involved generating the same amount of data as used in the fitting from a specific set of estimated parameters and then performing a parameter estimation on those data. Since the true parameter values for the generated data are known, we will confirm that the estimation converges to the true value. By computing the information gain in bit/fixation (Kümmerer et al., 2015). These robustness checks have previously been successfully run on other data. Additionally, we have run test estimations with the same amount of data as we will have in the present study. Therefore if these robustness checks fail, this is an indication that either the data are too noisy or that some aspect of the data fails to constrain the model. If the latter is the case, we will consider reducing the number of parameters estimated by fixing some parameter values. In both cases this choice will be reported and interpreted in the final paper.

C.9 Analysis plan

3.9.1 Statistical analyses

Testing H1: We will fit the model to the different tasks and compare the parameter values for σ_A, the parameter that controls the attentional span in the model. We will also use the model fits to generate data and compare the mean saccade amplitude of the empirical and simulated data.

Testing H2: We will fit the model to the different tasks and compare the parameter values for σ_I, a parameter that relates to the strength of the inhibition stream. We will also use the model fits to generate data and compare the mean lag distance of the empirical and simulated data.

Testing H3: We will fit the model to the different tasks and compare the parameter values for ω, the parameter that controls the speed of information decay in the model. Additionally, we will evaluate the model likelihood on the test data using a group of modified models. The modification adapts the SceneWalk model to be a Markov process of a fixed order n, where n is in the range $(1, 10)$. The generated priority maps for fixation selection will include information from exactly n past fixations. If the model likelihood improves when n is increased, this indicates that more past information is relevant for predicting the next fixation. We expect, if H3 is true, for the model likelihood to peak at a lower value of n for the free-viewing-like tasks than the search-like tasks.

3.9.2 Other analyses

Given the estimated model parameters we will simulate data. We will then compare the tendencies in the simulated data to the empirical data using the analyses reported in https://arxiv.org/abs/1911.06085, https://doi.org/10.1167/jov.0.0.06824

3.9.3 Exploratory analyses

In previous analyses the parameter CF could not be fitted in the same estimation procedure as the other parameters, since it destabilized the model. Here we will try to estimate CF separately, in a second estimation after the other parameters have been estimated. After fitting the parameters for the four tasks, we will use the parameter's posteriors as priors for a by-subject parameter estimation, implementing a kind of hierarchical fitting procedure. Estimating parameters for each subject and task will otherwise be difficult on the given data set, because there is likely not enough data to constrain the model in the training set or to evaluate it on the test set.

D Appendix for Paper 3

D.1 Discretization

The model is defined on a 100×100 lattice. In order to evaluate experimental eye movement traces (which are typically given in degrees of visual angle), we discretized the data. Each data point was multiplied by 350 and then applying the floor function. This discretization value for the entire data set was chosen by visual inspection, as it allowed all eye movement traces to stay within the confines of the grid, but also efficiently used the space. Parameter values from the stepping distribution to the potential critically depend on the specifics of this discretization.

D.2 Simulated data

Figure D.1 shows some examples of experimental gaze trajectories which illustrate that the model captures individual differences in the data that can be validated by visual inspection.

D.3 Parameter recovery

Parameter recovery analyses are an important step in evaluating the stability and reliability of a mathematical model. This is often done as a first step in model building to ensure that the model is able to accurately capture the underlying dynamics of the system being studied. Parameter recovery analyses involves generating synthetic data with known ("true") parameter values of the model and then estimating the parameters from this simulated data. This procedure permits the evaluation of the accuracy and precision of the parameter estimates, as well as the sensitivity of the estimates to the choice of estimation method and the quality and amount of data. It is an important step to assess the overall performance of the model and identify potential issues that may arise in the estimation process.

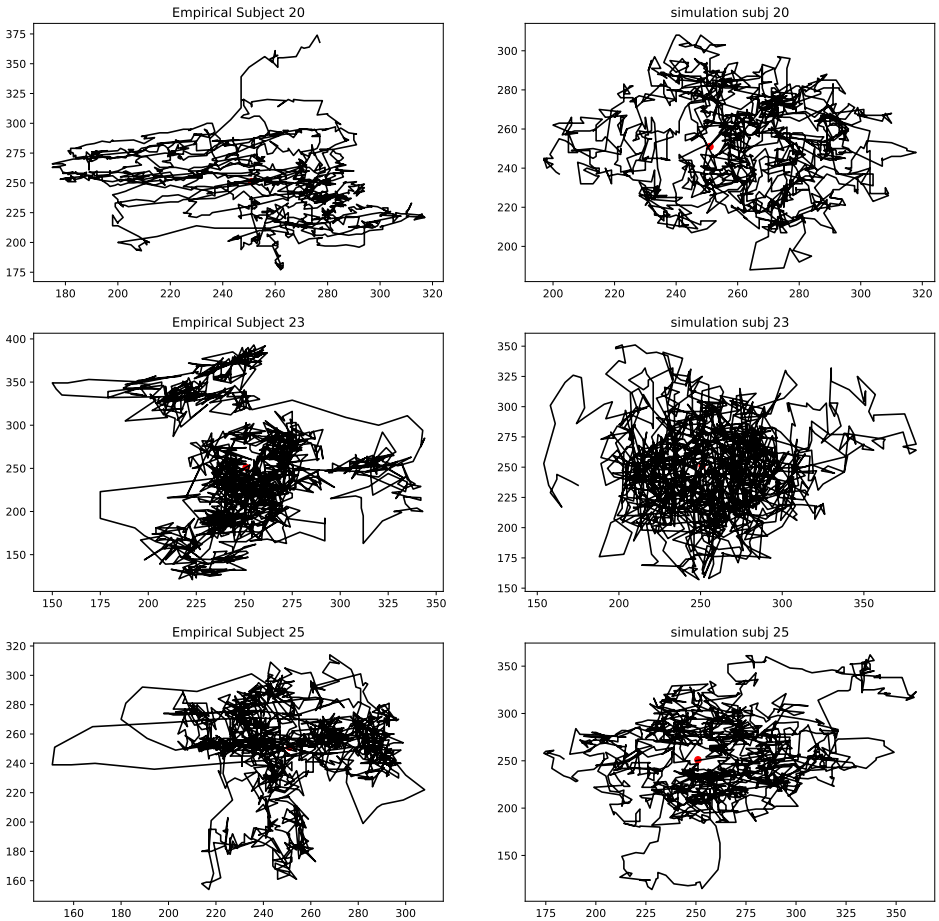

Figure D.1 Some examples of fixational eye movement traces. The left column shows the experimental data. The right column shows data simulated using the individually fitted models. These examples qualitatively illustrate the captured statistical properties of the drift movements.

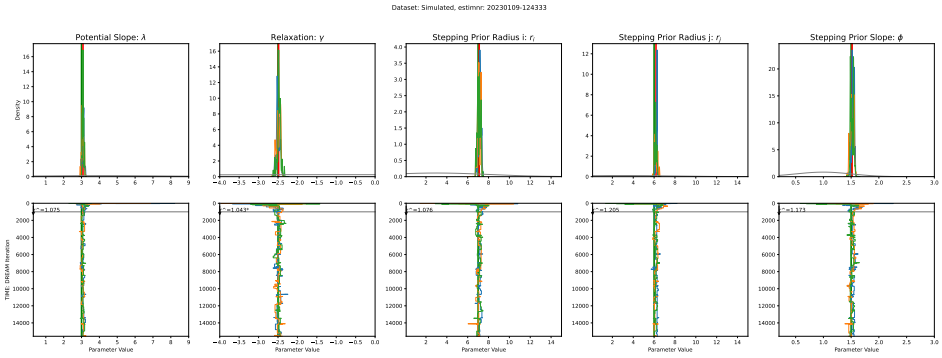

Figure D.2 Parameter Recovery Analysis. For each parameter we show that a recovery analysis converges to the correct value. The top row shows the posteriors relative to the priors and reveals strong convergence in all parameters. The bottom row shows the caterpillar plots of the estimation procedure, i.e., the parameter value that the three chains assumed in each iteration. The horizontal line indicated the burn in period.

D.4 Priors

For the parameter estimation we chose truncated Gaussian priors. Table B1 details the numerical values that define the truncated Gaussian priors.

	Potential Slope: λ	Relaxation: γ	Stepping Prior Radius i: r_i	Stepping Prior Radius j: r_j	Stepping Prior Slope: ϕ
mean	2	-2.5	5	5	1.5
sd	2	0.3	5	5	0.5
lower	0.3	-3.8	0.1	0.1	0.5
upper	8	0	15	15	3

Table B1 The parameters that define the prior distributions used during parameter inference. For each parameter we report the mean, standard deviation, as well as the upper and lower bounds of the truncated Gaussians.

D.5 Parameter estimation results

Table B2 gives the detailed point estimates for all estimated parameters for all participants, as well as 98% confidence intervals. Note that the participant IDs start at 20, since we estimated parameters for the final model for participant IDs 20 to 39 only. The data for participant IDs 1 to 19 were used for model building and are omitted here to prevent overfitting.

Subject	Potential Slope: λ		Relaxation: γ		Stepping Prior Radius i: r_i		Stepping Prior Radius j: r_j		Stepping Prior Slope: ϕ	
	mean	+/-	mean	+/-	mean	+/-	mean	+/-	mean	+/-
20	4.969	0.194	-3.585	0.209	3.304	0.124	2.664	0.083	1.080	0.021
21	4.497	0.162	-3.688	0.110	3.850	0.204	3.378	0.166	1.072	0.032
22	4.917	0.185	-3.727	0.072	3.153	0.106	2.333	0.081	1.062	0.028
23	4.533	0.045	-3.773	0.027	12.06	0.572	8.181	0.376	1.203	0.055
24	4.796	0.179	-3.693	0.104	3.402	0.152	2.644	0.091	1.077	0.030
25	4.563	0.167	-3.757	0.043	5.156	0.254	4.172	0.215	1.010	0.032
26	4.646	0.144	-3.767	0.033	7.301	0.266	4.562	0.195	1.199	0.026
27	5.100	0.163	-3.692	0.105	2.766	0.143	2.171	0.109	0.963	0.031
28	4.940	0.165	-3.700	0.098	3.558	0.171	2.152	0.071	1.052	0.035
29	5.211	0.165	-3.723	0.076	3.353	0.093	2.700	0.077	1.111	0.022
30	4.778	0.217	-3.750	0.050	3.700	0.133	2.193	0.104	0.972	0.028
31	4.621	0.115	-3.771	0.028	7.844	0.296	5.613	0.209	1.191	0.031
32	4.504	0.146	-3.759	0.041	11.30	0.393	6.701	0.223	1.217	0.039
33	4.565	0.223	-3.696	0.103	3.527	0.157	2.499	0.102	1.012	0.028
34	4.922	0.160	-3.720	0.080	4.331	0.207	3.204	0.165	1.085	0.037
35	4.693	0.209	-3.736	0.061	3.418	0.125	2.502	0.087	1.005	0.021
36	4.708	0.132	-3.765	0.035	7.780	0.336	5.351	0.220	1.220	0.037
37	4.223	0.104	-3.755	0.044	8.409	0.286	6.516	0.222	1.211	0.030
38	4.868	0.142	-3.732	0.067	3.914	0.116	2.751	0.091	1.092	0.025
39	5.061	0.091	-3.712	0.084	4.460	0.188	3.176	0.128	1.163	0.035

Table B2 Point estimates and 98% confidence intervals of the 5 estimated parameters for each subject.

E Preregistration for DAEMONS

E.1 Study information

5.1.1 Title

Potsdam **Da**taset for **E**ye **M**ovement **O**n **N**atural **S**cenes (Potsdam DAEMONS)

5.1.2 Authors

Lisa Schwetlick, Matthias Kümmerer, Matthias Bethge, Ralf Engbert

5.1.3 Description

In this study we will collect experimental data from participants viewing natural scenes. We want to generate a data set that is useful for both machine learning applications as well as experimental analysis. Previous data sets from the field of machine learning, such as MIT1003 (Judd et al., 2009) or CAT200 (Borji & Itti, 2015), encompass a large number of images seen by a comparatively small number of subjects for a fairly short amount of time. The eye tracking recordings are often only available only as ordered scan paths, sometimes omiting fixation durations altogether (Borji & Itti, 2015). By contrast, data sets from experimental eye movement research tend to have more subjects and longer fixation sequences but fewer images: having more data per image allows to quantify effects on a per-image bases more precisely (e.g., the SpatStat data set (Trukenbrod et al., 2019)). This leads to a comparability problem: machine learning models, such as Deepgaze III (Kümmerer & Bethge, 2021), which need large amounts of data with many different image examples can not be trained on the same data as the procedural models such as SceneWalk (Schwetlick, Rothkegel, Trukenbrod, et al., 2020b), which require longer sequences and vice versa. Additionally for evaluation purposes, its often good to have many observers per image. Therefore, since mechanistic models need less training data, vision science traditionally chooses a different tradeoff: many subjects on fewer images. For fair comparison and insightful evaluation we propose to collect a single data set that combines both: Many images with potentially fewer subjects per image as a training set, and many subjects per image for insightful evaluations as a validation and test set.

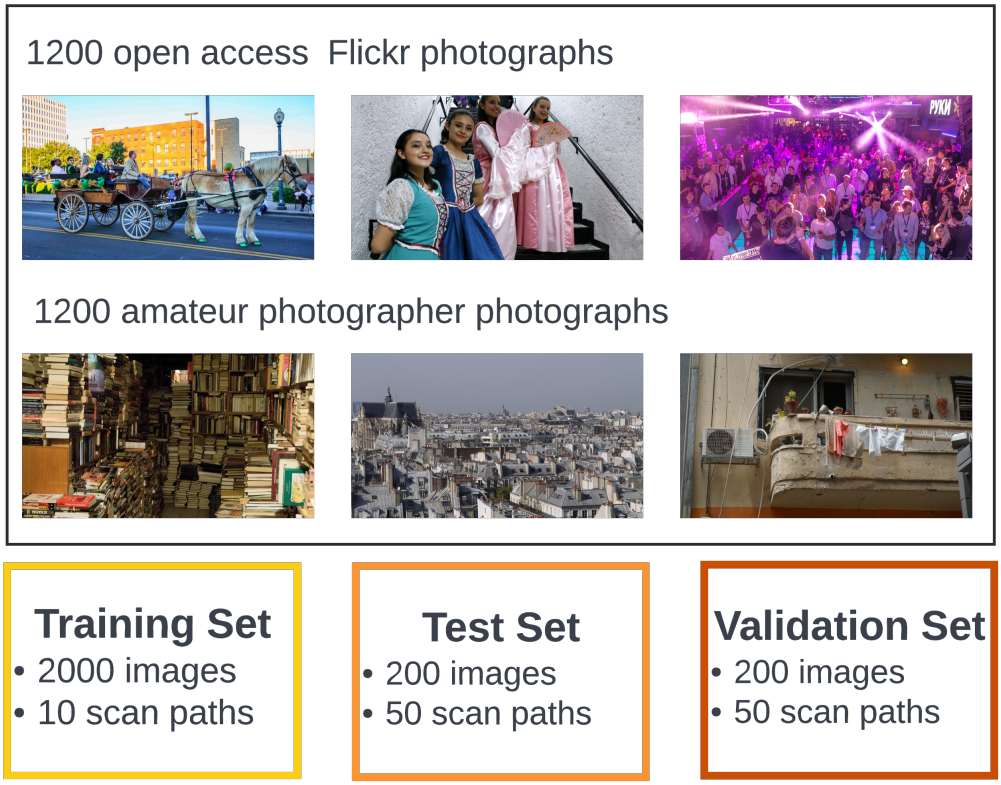

Figure E.1 The image data set for the DAEMONS study consists of 2400 photographs. One half was collected from amateur photographers and is unique to this data set. The other half are open access photographs from the platform Flickr. The images are subdivided into training, test, and validation data subsets.

We aim to address this gap by collecting a data set which can be used by both research traditions, and which will enable us to combine and compare insights from research of the visual system with state of the art machine learning techniques. It will also be invaluable in establishing a benchmark (see Kümmerer et al., 2018 for static saliency benchmark) for scan path modeling.

5.1.4 Hypotheses

This is purely a data publication. Our aims are chiefly to establish a useful data set on which to train models of eye movement.

E.2 Design plan

5.2.1 Study type

Observational Study

5.2.2 Blinding

No blinding is involved in this study.

5.2.3 Is there any additional blinding in this study?

No.

5.2.4 Stimulus material

For this study we will have a

- Training set (will comprise a large number of images seen by a smaller number of people)
- Validation set (will be a smaller number of images seen by all people)

The setup will be the following:

- total # training images = 2000
- total # images in validation set = 200
- total # images in test set = 200
- total # images each subject sees = 160
- total # times each training image is seen = 10
- total # times each test/validation image is seen = 50
- total # training images each subject sees = 80
- total # of subjects = 250

We will collect an image data set especially for this study. We initially considered using an existing image data set which already has a number of semantic labels already attached to the images. However all existing data sets we found had one or more of the following issues:

- Insufficient resolution: for a scene viewing data set that holds up to the standards of experimental cognitive science, it is important that the images are presented at large resolution (Otero-Millan et al., 2013; von Wartburg et al., 2007)
- Unnatural image material: in order to elicit scene viewing behavior it is important that images do not include blurred out areas (for example when faces are blurred for privacy) as is the case, for example, in the mapillary data set (https://www.mapillary.com/dataset/vistas). Also images with large amounts of text should occur at most rarely since they would alter the scene viewing behaviour.

Thus, we concluded we would have to build our own image data set. We selected the photographs according to the following criteria:

- minimal size of 1920x1080px
- landscape format
- try to minimize central/photographers bias, unfocused regions (such as is frequent in portraits) and writing
- try to maximize variability both between and within images.

One half of the images are freely available creative commons images taken from the platform *Flickr* (https://flickr.com). The other half are images taken by photographers, where we paid photographers to take pictures and release them into creative commons. All 2200 images will be published as a coherent data set on Open Science Framework (https://osf.io).

5.2.5 Study design

250 subjects with normal or corrected to normal vision will participate in the study. Their eye movements will be recorded as they look at 160 images from the stimulus data set described above. Each image will be shown for 8 seconds. Participants will be instructed to blink as little as possible and to carefully investigate the images. After every 20 images participants will be given a recognition task, where they will be shown 3 images and have to chose the unknown image between 2 previously seen images. This is done to ensure attentive participation. Each correct answer will gain the participant points. At the end of the experiment they will be paid an additional sum of money according to the number of points they gathered. A calibration of the eye tracker will be done every 20 trials.

5.2.6 Randomization

The images will be randomized over subjects and the training and validation set are mixed together. No subject will see any single image twice.

E.3 Sampling plan

5.3.1 Existing data

There is no pre-existing data

5.3.2 Explanation of existing data

None

5.3.3 Data collection procedures

We will collect data using a Eyelink 1000 eye tracker (1000Hz) monocular. The participants will have normal or corrected to normal vision and will be paid in either money or study credits. The data collection will start with a calibration of the eye tracker which will be repeated every 20 trials to ensure good quality of the data. We will use binocular tracking to measure the trajectories of both eyes. The images will be presented at the maximum trackable size for the Eyelink 1000 Desktop mount (32°visual angle, 95cm distance to the monitor). Each image will be presented for 8 seconds.

5.3.4 Sample size

For the training set we will collect 10 000 fixation sequences on 1000 images. Each image will be seen by 10 subjects. Given a viewing time of 8 seconds, each fixation sequence will be between 5 and 15 fixations long.

5.3.5 Sample size rationale

Author M.K. advised about the number of images needed to train a model like (Kümmerer & Bethge, 2021). Author L.S. developed a design that is feasible from an experimental perspective. No power analyses were conducted because no single effect is being investigated, however the data set will be larger than most in the literature and therefore we expect it to be sufficient also to conduct analyses regarding specific effects.

E.4 Variables

5.4.1 Measured variables

We will be measuring the coordinates of the eye position in the image at every millisecond. This raw data will be converted into fixations and saccades with the appropriate measures e.g., fixation duration, saccade amplitude and speed.

E.5 Analysis plan

5.5.1 Data exclusion

Trials during which the subjects do not participate well in the task, e.g., make no eye movements or close their eyes, will be excluded.

5.5.2 Missing data

Data missing due to blinks will be interpolated where possible and data points removed if necessary.

F A Brief Study on Writing

Das Werk ist die Totenmaske der Konzeption.

Walter Benjamin

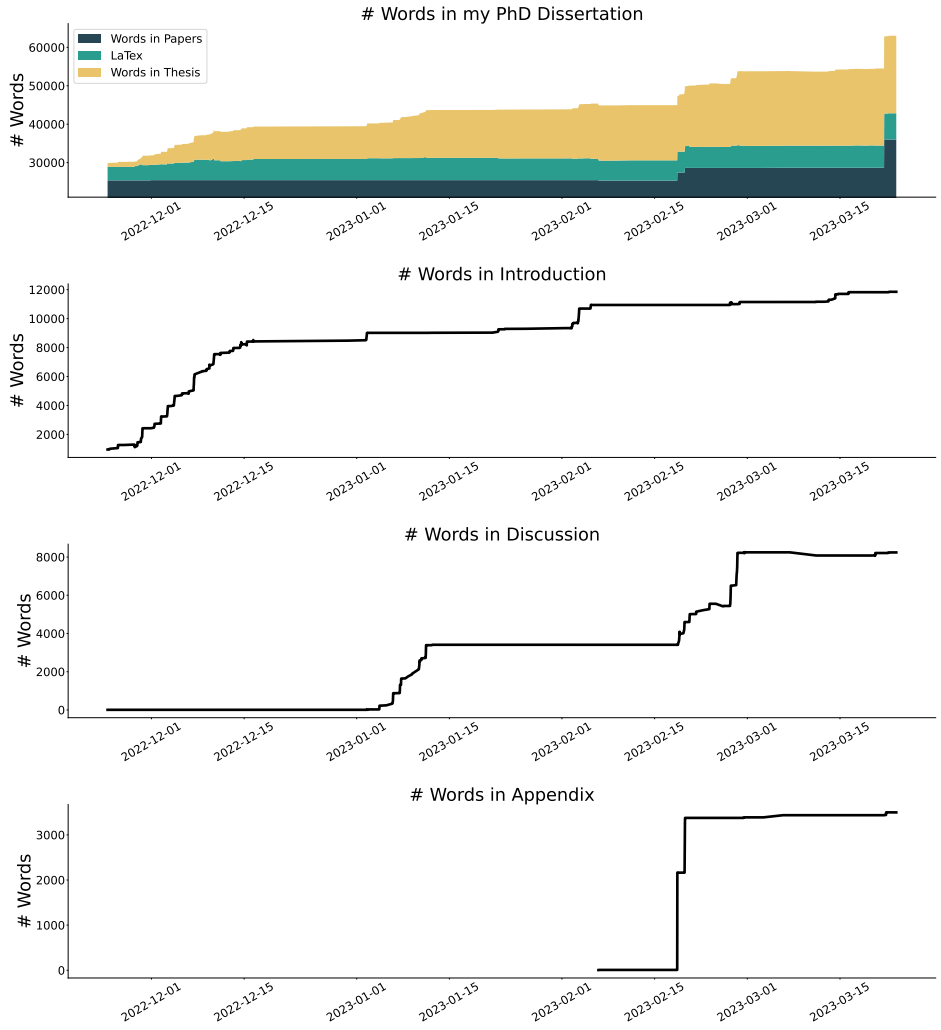

Figure F.1 Number of words in this manuscript over time, by section. The author (yours truly) experienced alternating phases of high output, editing, and reduced productivity. Note that the writing of paper #3 occurred in parallel to the writing of this text.

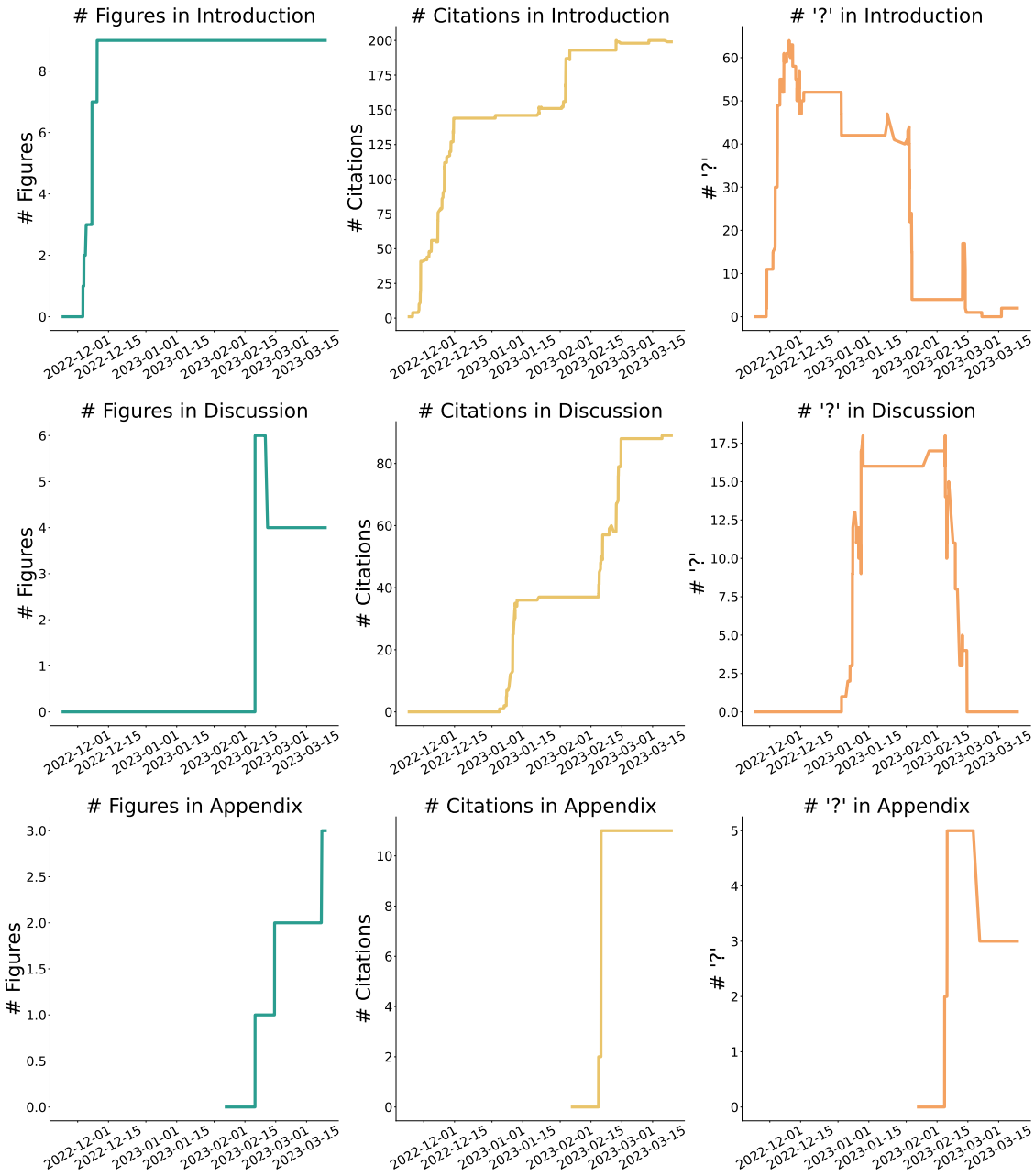

Figure F.2 Further count metrics, over time, by section. Rows represent the sections; columns represent the number of figures, citations, and question marks in the text, respectively.

Publications

Full First Author Journal Papers

Schwetlick, L., Backhaus, D., & Engbert, R. (2022a). A dynamical scan-path model for task-dependence during scene viewing. *Psychological Review*, *130*(3), 807–8. https://doi.org/10.1037/rev0000379

Schwetlick, L., Reich, S., & Engbert, R. (2023). *Bayesian dynamical modeling of fixational eye movements* [Preprint]. ArXiv. https://doi.org/10.48550/arXiv.2303.11941

Schwetlick, L., Rothkegel, L. O. M., Trukenbrod, H. A., & Engbert, R. (2020b). Modeling the effects of perisaccadic attention on gaze statistics during scene viewing. *Communications Biology*, *3*(727), 1–11. https://doi.org/10.1038/s42003-020-01429-8

Conference Contributions

Schwetlick, L., Backhaus, D., & Engbert, R. (2022b). Modeling task-dependency of eye movement during scene viewing. In V. McGowan, A. Pagán, K. B. Paterson, D. Souto, & R. Groner (Eds.), *Book of abstracts of the 21st european conference on eye movements*. Journal of Eye Movement Research, 15 *(5)*. https://doi.org/10.16910/jemr.15.5.1

Schwetlick, L., Kümmerer, M., Engbert, R., & Bethge, M. (2022). DeepGaze vs SceneWalk: What can DNNs and biological scan path models teach each other? *Journal of Vision*, *22(14):3986*. https://doi.org/10.1167/jov.22.14.3986

Schwetlick, L., Rothkegel, L. O. M., & Engbert, R. (2019). Adding neurally-inspired mechanisms to the SceneWalk model improves scan path predictions for natural images. *2019 Conference on Cognitive Computational Neuroscience*. https://doi.org/10.32470/ccn.2019.1206-0

Schwetlick, L., Rothkegel, L. O. M., & Engbert, R. (2020). Peri-saccadic attention drives saccade statistics in scene viewing. *Journal of Vision*, *20(11):700*. https://doi.org/10.1167/jov.20.11.700

Schwetlick, L., Rothkegel, L. O. M., Trukenbrod, H. A., & Engbert, R. (2017). Central fixation bias: The role of sudden image onset and early gist extraction. *European Conference on Visual Perception 2017*. https://doi.org/10.16910/jemr.10.6.1

Schwetlick, L., Trukenbrod, H. A., & Engbert, R. (2018). The influence of visual long term memory on eye movements during scene viewing. *European Conference on Visual Perception 2018*.

Schwetlick, L., Trukenbrod, H. A., & Engbert, R. (2019). The effect of visual long-term memory on eye movements over time. *Journal of Vision*, *19(10):149a*. https://doi.org/10.1167/19.10.149a

Preregistrations

Schwetlick, L., Backhaus, D., Brunken, R., & Engbert, R. (2022). *The effect of illumination-level on measurement stability using an eyelink1000 eye tracker* [Preregistration]. https://doi.org/10.17605/OSF.IO/3GUK4

Schwetlick, L., Backhaus, D., & Engbert, R. (2020). *Modelling advanced natural tasks using scenewalk* [Preregistration]. OSF. https://osf.io/dsyt2/

Schwetlick, L., Kümmerer, M., Bethge, M., & Engbert, R. (2022). *Potsdam dataset for eye movement on natural scenes (potsdam daemons)* [Preregistration]. https://doi.org/10.17605/OSF.IO/BDXGS

Data Publications

Schwetlick, L., Backhaus, D., Trukenbrod, H. A., & Engbert, R. (2020). *"Memory": Image familiarity and eye movement* [Data set]. Open Science Framework. https://doi.org/10.17605/OSF.IO/E7FVP

Trukenbrod, H. A., Schwetlick, L., & Engbert, R. (2020). *Spatial statistics for gaze patterns of repeated viewing in scene perception* [Data Set]. Open Science Framework. https://doi.org/10.17605/OSF.IO/ME2SH

Other Contributions

Engbert, R., Rabe, M. M., Schwetlick, L., Seelig, S. A., Reich, S., & Vasishth, S. (2022). Data assimilation in dynamical cognitive science. *Trends in Cognitive Sciences*, *26*(2), 99–102. https://doi.org/10.1016/j.tics.2021.11.006

Makowski, S., Jäger, L. A., Schwetlick, L., Trukenbrod, H. A., Engbert, R., & Scheffer, T. (2020). Discriminative viewer identification using generative models of eye gaze. *Procedia Computer Science*, *176*, 1348–1357. https://doi.org/10.1016/j.procs.2020.09.144

Malem-Shinitski, N., Opper, M., Reich, S., Schwetlick, L., Seelig, S. A., & Engbert, R. (2020). A mathematical model of local and global attention in natural scene viewing. *PLoS Computational Biology*, *16*(12), 1–21. https://doi.org/10.1371/journal.pcbi.1007880

Open Source Software

- SceneWalk https://github.com/lschwetlick/SceneWalk_Model
- SAW https://github.com/lschwetlick/sawpy
- Image Annotation https://github.com/lschwetlick/ImageAnnotationTool

- Saccade Detection https://github.com/lschwetlick/EngbertMicrosaccadeToolbox

Declaration of Authorship

Eigenständigkeitserklärung

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken (dazu zählen auch Internetquellen) entnommen sind, wurden unter Angabe der Quelle kenntlich gemacht.

Declaration of authorship

I hereby certify that the thesis I am submitting is entirely my own original work except where otherwise indicated. I am aware of the University's regulations concerning plagiarism, including those regulations concerning disciplinary actions that may result from plagiarism. Any use of the works by any other author, in any form, is properly acknowledged at their point of use.

23.03.2023, Berlin.

Lisa Schwetlick