Parafoveal processing of lexical information during reading

From experiments to computational modeling

Stefan A. Seelig

Doctoral Thesis submitted to the Faculty of Human Sciences at the University of Potsdam (Department of Psychology, Experimental and Biological Psychology) in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Potsdam, December 2020

Reviewers:

Reviewer 1:
Prof. Ralf Engbert
Allgemeine und Biologische Psychologie, Universität Potsdam

Reviewer 2:
Prof. Ronan G. Reilly
Computer Science, Maynooth University

Date of defense:
May 21, 2021

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-50874
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-508743

Abstract

During sentence reading the eyes quickly jump from word to word to sample visual information with the high acuity of the fovea. Lexical properties of the currently fixated word are known to affect the duration of the fixation, reflecting an interaction of word processing with oculomotor planning. While low level properties of words in the parafovea can likewise affect the current fixation duration, results concerning the influence of lexical properties have been ambiguous (Drieghe, Rayner, & Pollatsek, 2008; Kliegl, Nuthmann, & Engbert, 2006). Experimental investigations of such lexical parafoveal-on-foveal effects using the boundary paradigm have instead shown, that lexical properties of parafoveal previews affect fixation durations on the upcoming target words (Risse & Kliegl, 2014). However, the results were potentially confounded with effects of preview validity.

The notion of parafoveal processing of lexical information challenges extant models of eye movements during reading. Models containing serial word processing assumptions have trouble explaining such effects, as they usually couple successful word processing to saccade planning, resulting in skipping of the parafoveal word. Although models with parallel word processing are less restricted, in the SWIFT model (Engbert, Longtin, & Kliegl, 2002) only processing of the foveal word can directly influence the saccade latency.

Here we combine the results of a boundary experiment (Chapter 2) with a predictive modeling approach using the SWIFT model, where we explore mechanisms of parafoveal inhibition in a simulation study (Chapter 4). We construct a likelihood function for the SWIFT model (Chapter 3) and utilize the experimental data in a Bayesian approach to parameter estimation (Chapter 3 & 4).

The experimental results show a substantial effect of parafoveal preview frequency on fixation durations on the target word, which can be clearly distinguished from the effect of preview validity. Using the eye movement data from the participants, we demonstrate the feasibility of the Bayesian approach even for a small set of estimated parameters, by comparing summary statistics of experimental and simulated data. Finally, we can show that the SWIFT model can account for the lexical preview effects, when a mechanism for parafoveal inhibition is added. The effects of preview validity were modeled best, when processing dependent saccade cancellation was added for invalid trials. In the simulation study only the control condition of the experiment was used for parameter estimation, allowing for cross validation. Simultaneously the number of free parameters was increased. High correlations of summary statistics demonstrate the capabilities of the parameter estimation approach. Taken together, the results advocate for a better integration of experimental data into computational modeling via parameter estimation.

Zusammenfassung

Während des Lesens springt der Blick von Wort zu Wort, um visuelle Informationen mithilfe der hohen Auflösung der Fovea aufzunehmen. Lexikalische Eigenschaften eines zurzeit fixierten Wortes wirken sich dabei auf die Fixationsdauer aus, was eine Interaktion von Wortverarbeitung mit okulomotorischer Bewegungsplanung impliziert. Während Low-Level-Eigenschaften eines parafovealen Wortes ebenfalls die Fixationsdauer beeinflussen können, sind Ergebnisse zu Einflüssen lexikalischer Eigenschaften parafoveler Worte uneindeutig (Drieghe et al., 2008; Kliegl et al., 2006). Experimentelle Untersuchungen solcher parafoveal-on-foveal-Effekte mittels des Boundary-Paradigmas zeigten stattdessen, dass sich lexikalische Eigenschaften parafovealer Worte auf Fixationsdauern auf den Target-Wörtern auswirken (Risse & Kliegl, 2014). Diese Ergebnisse waren jedoch möglicherweise mit den Effekten der Preview-Validität konfundiert.

Die Möglichkeit parafovealer Verarbeitung lexikalischer Informationen stellt bestehende Modelle für Blickbewegungen beim Lesen vor Probleme. Modelle, die auf seriellen Wortverarbeitungsannahmen fußen, können derlei Effekte nicht schlüssig erklären, da in ihnen erfolgreiche Wortverarbeitung oft starr an Bewegungsplanung gekoppelt ist, was ein Überspringen des parafovealen Wortes zur Folge hätte. Obwohl Modelle mit paralleler Wortverarbeitung weniger eingeschränkt sind, kann im SWIFT-Modell (Engbert et al., 2002) nur die Verarbeitung fovealer Worte die Sakkadenplanung direkt hemmen.

Wir verbinden in dieser Arbeit die Ergebnisse eines Boundary-Experiments (Kapitel 2) mit einem prädiktiven Modellierungsansatz mit dem SWIFT-Modell, in dem wir Mechanismen parafovealer Hemmung in einer Simulationsstudie erkunden (Kapitel 4). Wir konstruieren eine Likelihood-Funktion für das SWIFT-Modell und nutzen die Experimentaldaten in einem Bayesianischen Ansatz zur Parameterschätzung (Kapitel 3 & 4).

In den Ergebnissen des Experiments zeigt sich ein substanzieller Frequenzeffekt des Previews auf die Fixationsdauer auf dem Target-Wort, der klar vom Effekt der Preview-Validität unterschieden werden kann. Mittels der Blickbewegungsdaten der Probanden demonstrieren wir die Praktikabilität des gewählten Ansatzes selbst mit nur wenigen freien Parametern, indem wir Statistiken der Probanden mit jenen aus Simulationen auf der Basis geschätzter Parameter vergleichen. Schließlich können wir zeigen, dass SWIFT die lexikalischen Preview-Effekte erzeugen kann, wenn das Modell zusätzlich mit einem Mechanismus parafovealer Inhibition ausgestattet wird. Die Effekte der Preview-Validität wurden hingegen am besten modelliert, wenn eine Möglichkeit zum Abbruch der Sakkadenplanung in Abhängigkeit von der Wortverarbeitung hinzugefügt wurde. In dieser Simulationsstudie wurden lediglich Daten der Kontrollbedingung des Experiments zur Parameterschätzung genutzt, wodurch eine Kreuzvalidierung der Güte der Simulationsdaten ermöglicht wurde. Gleichzeitig wurde die Zahl der freien Parameter erhöht. Hohe Korrelationen der Statistiken verdeutlichen das Potential des Parameterschätzungsansatzes.

Zusammengenommen sprechen die Ergebnisse dafür, dass Experimentaldaten mehr zur computationalen Modellierung herangezogen werden sollten, indem Möglichkeiten der Parameterschätzung ausgenutzt werden.

Contents

List of Figures

List of Tables

Chapter 1

General introduction

Reading is a critical skill that permeates every aspect of modern life. It depends on the seamless interplay of visual and linguistic processing with oculomotor coordination, and has therefore sparked the interest of cognitive scientists and pioneers of eye movement research alike. According to Rayner (2009b), four eras of research on eye movements during reading can be distinguished.

The first era begins with Javal's observations (by 1879, Wade & Tatler, 2008) of the well-known motion patterns during reading, and was concerned with outlining the research area by establishing basic facts and vocabulary. During the second era, interest in eye movements during reading declined due to technical limitations, as well as the behaviorist reservations towards inferring mental processes. Improved theories of language, accompanied by technological advances, marked the third era. They provided more accurate measurements and gave rise to new experimental paradigms, specifically gaze contingent techniques, which allowed for experimental configurations to react to eye movements almost instantly.

Finally, the fourth era coincides with the present day. It is characterized by the development of sophisticated computational models, whose predictions and constraints motivate a considerable amount of research. While sophisticated models are not uncommon in the general field of cognitive psychology, extant models of eye movements during reading stand out because of their complexity and realism. They are capable of generating a wide spectrum of behavior in the same format as experimental data, affording easy comparisons of model predictions and laboratory measurements. Among those models, two main contenders implement opposing hypotheses on the order of word processing, and the connection between word processing and oculomotor control. In the E-Z Reader model (Reichle, Pollatsek, Fisher, & Rayner, 1998) words in a sentence are processed in a rigid, serial order. Eye movements are generated as a result of successful word processing. Conversely, in the SWIFT model (Engbert et al., 2002), several words can be processed in parallel, and eye movements are generated by an independent timing process. Rayner (2009b) further points to three controversial problems in empirical research, that exhibit discrepancies in the comparison between experimental and correlative studies: Word skipping, regressions and parafoveal-on-foveal effects.

In this thesis I present empirical evidence for parafoveal processing of lexical information, strongly related to lexical parafoveal-on-foveal effects, in an experimental study, and explore mechanisms in the framework of the SWIFT model via numerical simulations, that can account for these effects. In this context, I describe and apply a method of parameter estimation for complex models, that allows to incorporate individual differences into the model, based solely on recorded eye movement data.

1.1 Eye movements during reading

Reading is not an evolved skill, but a cultural accomplishment that builds on the interplay of brain functions like learning, abstraction, pattern recognition and oculomotor coordination. To understand a sentence, all words must be processed and represented on different levels of analysis: First, letter identities and word length form a word's *orthographical* representation, after which its acoustical and basic grammatical properties can be decoded, resulting in *phonological* and *morphological* representations. The "abstract representation of the word form" (i.e., how a word is written; Schotter, Angele, & Rayner, 2012) constitutes a word's *lexical* quality. Finally, word meaning and grammatical role are decoded, forming *semantic* and *syntactic* representations. How is it possible for our brain to accomplish such intricate decoding during what we perceive as merely brief glances? To gain a better understanding of processes involved in reading, it is useful to begin with a reflection on the role of the biology of vision.

1.1.1 Vision

The visual system is the single most important perceptual modality in humans. Processing visual information takes up more than a quarter of the human cortex (Van Essen, 2003). Vision plays an important role in the regulation of circadian rhythms, and constitutes our main route of assessing our environment.

The eyes are the sensory organ of this system. When light enters our eyes, it is projected through the lens onto the retina, where it is translated into neuronal code by two types of photoreceptors: Cone cells, responsible for photopic vision (i.e., vision under daylight conditions), and rod cells, responsible for scotopic vision (i.e., vision under low light conditions). The receptor types exhibit differences in their excitability by different wavelengths, degrees of connectedness to ganglion cells, as well as distribution patterns on the retina. Rod cells react to lower light intensities as cone cells and while many rod cells can converge on a single ganglion cell which adds up all contributions to calculate a response, this number is much lower for cone cells. Additionally, where all rod cells generally react to the same wavelengths, three different types of cone cells are selective to different, specific wavelengths and thus enable color vision. This allows for a more detailed perception during photopic vision, which is further helped by a massive aggregation of cone cells in the center of the retina—the *fovea*.

The fovea is an indented, circular area in the center of the retina of about $0.33cm$, or 2° visual angle, in diameter, where visual resolution is highest, and rod cells are completely absent. Outside of the fovea the number of cone cells drops rapidly, while the density of rod cells sharply rises (Remington, 2012). The annulus between the fovea and a concentric circle with a radius of 5° visual angle is called *parafovea*[1] (Rayner, 1998).

1.1.2 Eye movements

Due to the centralised organisation of the retina, only a small part of the visual field can be processed with high acuity at a time. During visual exploration, as well as reading, the gaze is therefore frequently relocated to new areas in the visual field with fast, ballistic movements called *saccades* about three to four times per second on average. Visual processing occurs only during the periods of stationarity between saccades, termed *fixations*, whereas during saccades the intake of visual information is actively suppressed (Matin, 1974). Average saccade lengths and fixation durations show considerable task dependent variation (Table 1, Rayner, 1998).

Saccadic movement is preceded by a phase of motor planning, which usually takes about 150-175ms (Rayner, 1998). Experiments have shown, that during this latency the saccade target can be updated until about $70ms$ before the onset of the saccadic motion (Becker & Jürgens, 1979), indicating that *saccade preparation* consists of at least two stages. In line with the account of separate, hierarchical stages is the finding, that fixation durations can be much shorter than 150ms, suggesting partial overlap, or parallel programming, of saccades (Morrison, 1984).

Other important types of eye movements are *smooth pursuit* when the eyes are entrained on the motion of a moving object, and *vergence movements*, when the point of fixation changes in distance towards or away from the observer, resulting in opposing movement directions of the eyes. The *vestibulo-ocular reflex* compensates for head movements and the *optokinetic reflex* returns the eyes to their original position, if the target of a smooth pursuit movement leaves the visual field.

During fixations—other than the name suggests—the eyes are not motionless, but slightly meander in small *drift movements*, overlaid by an oscillatory *tremor*. When the eyes drift too far from the original point of fixation, *microsaccades*, which share many characteristics with regular saccades but are involuntary and of small amplitudes (Martinez-Conde, Macknik, & Hubel, 2004), relocate the eyes to their initial fixation position. These

[1] In older publications, the parafovea is sometimes described as extending up to a radius of 10° (e.g., Ditchburn, 1971, according to Rayner, 1978). While anatomically the distinction is based on histological data (Remington, 2012), definitions vary in psychological literature.

fixational eye movements are seen to serve counteracting sensory adaptation, which, when enforced, results in perceptual fading (Ditchburn & Ginsborg, 1952).

1.1.3 Reading

During normal reading, saccades move the eyes through a sentence in the direction of the natural word order, with most words receiving at least one fixation. However, only about 50% of saccades are *forward saccades*, moving the gaze from a given word n the upcoming word $n+1$. During *skipping saccades* the eyes directly jump to the word $n+2$ or further in reading direction, whereas by *refixation saccades* the fixation position is slightly moved within the same word n. Both types each make up about 20% of saccades during reading. *Regressive saccades* move the eyes to a preceding word $n-1$ or further back, sometimes the beginning of the sentence, and account for the remaining 10%. Although the length of saccades is usually measured in degrees, letter spaces are the prevailing metric used in reading instead. Here, saccade sizes are mostly invariant when the same text is read at different distances, with 7 to 9 letter spaces being the typical saccade length (Rayner, 1998) in alphabetical scripts.

While it is generally assumed, that saccades aim roughly at word centers (Rayner, 1979), landing sites within words vary broadly. The landing site distributions can be explained as a linear combination of a preferred saccade length, the distance between launch site and the center of the target word, and the contribution of a random error (McConkie, Kerr, Reddix, and Zola, 1988; see also Krügel and Engbert, 2014). The stochastic components in the computation of landing sites can cause *mislocated fixations*, which can amount up to 30% of fixations (Engbert & Nuthmann, 2008). Mislocated fixations are often quickly followed by corrective saccades (Nuthmann, Engbert, & Kliegl, 2005), therefore reducing fixation durations for fixations close to word borders. This *inverted optimal viewing position* (IOVP) effect (Vitu, McConkie, Kerr, & O'Regan, 2001) stands in contrast to findings from single word presentation experiments, where the latencies in word naming tasks are lowest (i.e., fastest), when the word is fixated at the center (O'Regan & Jacobs, 1992), termed the *optimal viewing position* (OVP). The opposing observations prevalent in OVP and IOVP effects emphasize the importance of natural reading paradigms (i.e., sentence reading).

Foveal analysis plays a major role in successful reading. In word identification studies, parafoveally presented words are identified with less accuracy (Bouma, 1973), proportional to their eccentricity to the point of fixation. According to Just and Carpenter's 1980 *eye-mind* assumption, words are fixated as long as their processing continues, hence different measures of fixation durations can provide insight into the time course of word processing. Fixation durations are distinguished based on the number of consecutive fixations on a specific word and whether the word has been reached or passed in a previous sequence of progressive (i.e., forward or skipping) saccades. The *first pass* measures of *single fixation durations* (SFD) and *first fixation durations* (FFD) reflect processing during the first encounter with a word, whereas *gaze durations* (GD) are calculated as the sum of all first pass fixations, and therefore also reflects more complete word processing.

Fixation durations show considerable variation between, but also within individual readers, ranging from fixations shorter than $100ms$ to fixations outlasting $500ms$ (Rayner, 1998). In addition to contributions from motor uncertainties, this variance also reflects on-line word processing of *word frequency*, *word length*, and context based *predictability* of the fixated word. Further, previously previously fixated words, and adjacent, parafoveal words have also been shown to influence fixation durations (Kliegl et al., 2006).

The influence of word frequency on the duration of word processing can be most easily shown in lexical decision tasks, where subjects must quickly decide whether a presented string of letters is a word or a non-word. Reaction times correlate with word frequency, that is, the word's rate of occurrence in a representative language corpus. High frequency words elicit faster responses than words of low frequency (Gardner, Rothkopf, Lapan, & Lafferty, 1987). The rationale behind this effect is, that the participants' experience with the respective words influences the speed of their lexical decision. Similarly, word frequency is correlated with fixation durations during reading (Henderson & Ferreira, 1990).

1.1.4 Parafoveal processing

The observation that letters and words can be processed from parafoveal presentation (Bouma, 1973) lead to the development of *gaze contingent* experimental paradigms, which allow the systematic manipulation of parafoveal content during reading. They rely on technological advances that made it possible to record and process eye movements fast enough that the calculated positions can be directly used in experimental control architecture.

In the gaze contingent *moving window paradigm* (McConkie & Rayner, 1975), characters outside of a fixed range around the fixation location are replaced with $x$es, creating a tight window around the fixation position, outside of which no information can be extracted. This makes it possible to analyse the amount of visible text necessary for normal reading. A central finding using this paradigm, is that the *perceptual span* (i.e., the area around the fixation in which words can be processed) is asymmetric during reading. It extends about four letters back, and 14-15 letters in reading direction (e.g., to the right in english, but to the left in hebrew Pollatsek, Bolozky, Well, & Rayner, 1981). Inverting the window to mask foveal words while leaving parafoveal information intact drastically slows down reading, but does not eliminate successful reading comprehension (Rayner, Inhoff, Morrison, Slowiaczek, & Bertera, 1981). Together with the notion, that during reading some words are skipped and are never fixated, these findings suggest that word processing during reading significantly extends to parafoveal words.

The gaze contingent *boundary paradigm* (Rayner, 1975b) was developed to address specifically parafoveal processing of upcoming words. When subjects begin to read a sentence, it initially contains an altered version, the *preview*, of a specific *target word*. Type and extent of the alteration depend on the research question and can range from visual degradation, scrambled letters, or changes in capitalization, up to switching out the whole word. As the gaze progresses through the sentence, eventually it crosses an invisible boundary and thereby triggers a display change, during which the altered word is switched back to its unaltered form. The boundary is placed after the last letter of the word preceding the target word, so that the preview can not be fixated if the display change happens fast enough. This setup precludes any foveal analysis of the preview, but allows the reader to process it parafoveally. Unless it is set off unexpectedly due to drift movements during a fixation, the display change itself is not noticeable due to saccadic suppression (Matin, 1974).

Contrasting trials that contain alterations of the preview—so called *invalid previews*—to those that contain no manipulation—or *valid previews*—permits insights into early word processing before foveal analysis, as well as possible interactions with ongoing foveal processing of the previous word. Trials with invalid previews consistently yield longer fixation durations on the target word (i.e., after crossing the boundary), as compared to trials with valid previews, which is seen to indicate that the availability of a parafoveal preview facilitates the subsequent foveal processing of the target word. This *preview benefit* is found to encompass several, distinct levels of processing. For example, orthographic similarity affects the size of the preview benefit effect, when invalid previews with different letter overlap to the target word are compared (Balota, Pollatsek, & Rayner, 1985; Snell, Vitu, & Grainger, 2017). This implies, that the preview benefit effect generally reflects the amount of information shared between preview and target word (Rayner, Well, Pollatsek, & Bertera, 1982). Additionally, phonological (Pollatsek, Lesch, Morris, & Rayner, 1992), morphological (Deutsch, Frost, Pelleg, Pollatsek, & Rayner, 2003), and semantic similarities (Hohenstein & Kliegl, 2014; Hohenstein, Laubrock, & Kliegl, 2010; Yan, Richter, Shu, & Kliegl, 2009) were found to facilitate target word processing, demonstrating that parafoveal processing can extend to high levels of word representation.

Contrary to predictions of prevailing assumptions on word processing, corpus studies showed influences of lexical properties (i.e., as reflected in word frequency) of an upcoming word $n + 1$ on fixation durations on the preceding word n (Kliegl et al., 2006). The boundary paradigm was used to investigate such *parafoveal-on-foveal* (PoF) effects, believed to be the result of *cross-talk*—a phenomenon where properties of parafoveal words modulate fixation times on foveal words in experiments involving word pairs and word lists (Kennedy, 1998; Kennedy, Pynte, & Ducrot, 2002). However, during sentence reading only orthographical PoF effects were found consistently (Angele, Slattery, Yang, Kliegl, & Rayner, 2008; Inhoff, Starr, & Shindler, 2000), whereas lexical PoF effects remained spurious at best (see Brothers, Hoversten, & Traxler, 2017). However, an experiment using a preview manipulation two words after the boundary, on word $n + 2$, found, both, lexical PoF effects, as well as preview effects in fixation durations on word $n + 1$. Observing effects of lexical parafoveal processing in an upcoming fixation was assumed to be caused by a delay (Risse & Kliegl, 2012), which could also explain why PoF effects were found so inconsistently. Later investigations confirmed the existence of lexical preview effects in upcoming fixations (Risse & Kliegl, 2014). Their existence constitutes a major challenge for contemporary models of eye movements during reading.

1.2 Computational models

Computational modeling has become a central paradigm in cognitive psychology. The practice of implementing cognitive models as computer programs naturally leads to highly detailed, and highly specific formulations. Process oriented *computational models* aim to reproduce a phenomenon by simulating—to some degree of abstraction—the processes that are assumed to lie between stimulus and observation. Being an instantiation of an underlying theory, they increase the clarity of theoretical statements by forcing the modeler to make their hidden assumptions explicit (Fum, Del Missier, & Stocco, 2007). They often include complex forms of dynamic interactions with random or stochastic components that can otherwise be difficult to characterize as explicitly (Fum et al., 2007).

Computer code's amenability to in depth analysis makes exploring theories through computational models a valuable tool. Additionally when a cognitive model does not lend itself to closed form analysis, the computational approach offers a unique and convenient opportunity to simulate the outcome of experiments that permit evaluation congruent to experimental data. Where theory is vague or uncertain, competing ideas can be implemented and their effects evaluated and compared against the backdrop of the model, which in turn can give rise to testable predictions and inform future experiments.

A number of cognitive computational models of eye movements during reading have been proposed to examine various aspects in our understanding of the underlying relationships[2] (see Reichle & Schotter, 2020, for an exhaustive list). As an example, Legge, Klitz, and Tjan's (1997) ideal observer model *Mr. Chips* is aimed at simulating the constraints of clinically low vision. Among conceptualizations of the visual span, word processing and visuo-motor planning, it features a locally adjustable retinal resolution, endowing it with the ability to simulate scotomas and similar clinical impairments in order to study their effect on strategies of eye guidance. Given its limited scope, however, *Mr. Chips* lacks a temporal dimension and therefore makes no predictions regarding fixation durations. To provide a unifying account of experimental findings, comprehensive models of eye movements during normal sentence reading should explain the complex patterns of fixation durations and fixation positions in sequences of fixations.

[2]Although approaches to model eye movements during reading based on deep learning neural networks exist (Wang, Wang, & Wu, 2020), they are less motivated by theories of cognition and primarily aim at reproducing reading behavior, regardless of the transferability of their mechanisms.

Models of reading and eye movements can be compared along many dimensions. Historically, one dominant distinction refers to the amount of influence via cognitive processing on saccade generation. While *cognitive control* models assume strong connections between word processing and saccade generation, *oculomotor control* models postulate lower-level properties of visual circuitry as the primary underlying source of eye movements. A different distinction hinges on assumptions on attention and the order of word processing. *Serial attention shift* (SAS) models propose that words are processed in a strictly serial order and that only one word can be processed at a time. In opposition, *guidance by attentional gradient* (GAG), or *processing gradient* (PG) models assume that words within a graded area around fixation are processed in parallel center of fixation.

Historically, SAS models have turned out to be a stimulating approach for practical and theoretical research on reading. However, their inability to provide a simple framework that can elegantly explain the complete variety of eye movement behavior during reading has prompted a search for alternatives. PG models offer a novel approach to tackle similar problems. After 20 years of debate, two models largely dominate the debate between serial and parallel processing: The *E-Z Reader* model (Reichle et al., 1998), driven by sequential attention shifts, and the *SWIFT* model, using a dynamical and gradient based approach.

1.2.1 E-Z Reader

As model of the *cognitive control* class, Reichle et al.'s (1998) state-based automaton *E-Z Reader* emphasizes the role of cognitive processing as basis for the control of eye movements. All processes entertained by the model are represented as states, which, upon initiation, are assigned a fixed or stochastic duration specific to the state and other circumstances. Stochastic durations are sampled from gamma distributions. In its first form, the model contains three stages for oculomotor control, namely the labile and non-labile stages of saccade preparation and a concluding stage of saccade execution. Word processing is also comprised of three stages, the first of which being the obligatory deployment of attention, followed by the *familiarity check* (L1) and *lexical completion* (L2) stages of lexical processing. An interpretation of the two processing stages is borrowed from the dual-process theory of recognition (Reichle, 2011). *E-Z Reader*s conceptualizations of word processing and oculomotor control are rather simplistic. Although oculomotor and cognitive aspects of *E-Z Reader* have received several updates (Reichle, 2011), adding some complexity, the parsimony is intentional. The model aims to elegantly produce hallmark effects of reading with a minimal set of assumptions.

Sequential attention shifts (SAS) constitute a central design principle of the *E-Z Reader* model while also serving to classify it together with a number of similar models. SAS models assume a rigorous coupling of attention and word processing, which is largely restricted to a single word at a time. Another implication of SAS models is their requirement, that the words are processed in a serial order. Only when processing of a word has concluded, attention is shifted to the next word in reading direction, after which processing begins. Morrisons (1984) assumption that eye movements are indicative of attention shifts after word processing was successfully completed, served as inspiration for *E-Z Reader*. Different from Morrison's ideas, however, *E-Z Reader* posits two stages of lexical word processing, instead of one. Additionally, eye movement preparation is not triggered by attention shifts, but instead by the completion of the first stage of lexical processing. Thus eye movement preparation starts, before attention can be reallocated. Once stage L1 on a word n is completed, an eye movement to the upcoming word n_{+1} starts undergoing preparation, although attention is and will remain focused on word n, until also the second lexical processing stage L2 on word n is finished. In *E-Z Reader* any lexical processing of a word requires that attention is affixed to the word as well. In compliance with findings on saccadic dead time by Dutton and Starbuck (1971), saccade programming also consists of two distinct, hierarchical stages: A *labile motor program* (M1) that can be aborted in light of a newer targeting signal, and a *non-labile motor program* (M2), that cannot be stopped. The possibility of aborting a labile stage of saccade preparation enables word skipping: After completion of L1 on word n an eye movement is prepared towards word n_{+1}. If the L2 stage on word n is complete, attention shifts to word n_{+1}, initiating an L1 processing stage. If the L1 stage of lexical processing is completed earlier than the labile motor program M1, M1 is aborted and directed to word n_{+2}. Consequently, word n_{+1} will be skipped.

The cognitive and oculomotor subsystems of *E-Z Reader* can operate in parallel on their respective tasks without mutual interference. However, the oculomotor system critically depends on input from lexical processing, where completion of the L1 stage constitutes the only route to movement initiation via the oculomotor system, pertaining to the models cognitive focus.

After the M2 stage a saccade is performed to the center of the previously chosen word. Following the ideas of McConkie et al. (1988), later versions of the model also contain systematic and random saccade errors which perturb the landing site as a stochastic function of the the intended saccade length. Saccadic errors introduce more realistic landing site distributions, introduce a mechanism for refixations that the model previously lacked, and provide the basis for a saccadic correction mechanism as an additional source to regulate oculomotor control (Reichle, Rayner, & Pollatsek, 1999).

The early model was well received for providing a parsimonious, unifying mechanism that could account for effects of word frequency, word predictability, word lengths, parafoveal preview, interactions between parafoveal preview and foveal load, spillover effects and skipping costs. Over the years the model has stimulated plenty of research and received various updates to accommodate a wider array of observations. The addition of a third stage, *postlexical integration* (I), of word processing (Reichle, Warren, & McConnell, 2009) enables the model to perform regressive saccades. Although coupled to word processing—the third stage is started after completion of the regular L2 stage—it does not block attention from moving to the next word and can operate in parallel to regular lexical processing stages L1 and L2. Either if stage L2 on word n_{+1} finishes while there is an unfinished stage I on word n, or at random, with low probability, upon completion of stage I, a saccade program is triggered targeting word n where integration failed. This enables the model to perform regressive saccades to the preceding word.

1.2.2 SWIFT

The *processing gradient* model *SWIFT* of Engbert et al. (2002) was developed in response to perceived shortcomings of the SAS modeling approach. Despite several similarities—both models similarly respect the implications of research on temporal aspects of saccade preparation (Dutton & Starbuck, 1971), as well as results on errors in saccade targeting by McConkie et al. (1988) —fundamental differences arise as SWIFT replaces the SAS principle with one that strongly favors parallel, simultaneous word processing. Instead of both being triggered by word processing, saccade programming and target selection occur at different times, overall governed by a random timing process. In SWIFT, the parallel evolution of dynamically linked discrete random walks is the predominant design principle.

Words are processed based on their location within a processing gradient, which is centered and maximal at the fixation location, and decreases with eccentricity. A words processing speed is a reflection of the area of under the curve of the processing gradient that it occupies. At any time, each word is assigned an activation value pertaining to the progress of it's processing. In the beginning of the first stage of word processing, as well as the end of the second stage, a words activation values are lowest, whereas they are highest during the moment of transition between the two processing stages. The maximal height of a words activation is determined from its length and frequency, thus giving the model a route to incorporate

Because the height of the processing gradient is maximal at the center of fixation, words close to the fovea are processed faster than words in the periphery, i.e., their activations change more rapidly. Since the selection process for saccade targets favours words with high activations, quick processing of foveal words will reduce their activations, whereas slow processing of peripheral words increases their selection probability, as they approach the point of transition between the two processing stages. While the mechanism expressed in this relationship theoretically can be sufficient to elicit a selection bias in reading direction, the processing gradient also is asymmetric, resemblant of the perceptual span.

In a later addition, the spatial extent and curvature of the processing gradient were given dynamical properties as well. Akin to the *attentional zoom lens* model (Eriksen & James, 1986), it was coupled to the foveal word activation, extending when activation is low and contracting when activation is high.

The onset of saccade preparation is determined by an independent random timer. The timer receives inhibitory input that is dynamically calculated from the activation of the currently foveated word. High activation values delay the onset of saccade preparation. If the timer reaches completion it immediately restarts itself.

Saccade preparation is a hierarchical process that involves two consecutive random timers, representing the labile and non-labile stages of saccade preparation (Dutton & Starbuck, 1971), respectively. During the transition from the labile to the non-labile stage, a definitive saccade target is selected on the basis of the momentary word activation field. As in *E-Z Reader*, the labile stage can be aborted as a result of a new saccade program, whereas the non-labile stage cannot.

The *Dynamical field theory of movement preparation* (Erlhagen & Schöner, 2002) provides the framework of a dynamically evolving choice array, whose values serve as weights in a selection process. To select a saccade target, an activation field is formed from the activation values of all words. The selection probability of a word is proportional to its activation value in relation to the sum of all activations.

The executiion of a saccade briefly disrupts word processing and also is governed by a random timer in duration. Saccade landing sites are determined in agreement with the model of McConkie et al. (1988), with a systematic range error and oculomotor noise, both depending on the length of the intended saccade.

The SWIFT model is incredibly versatile and can account for many kinds of gaze behavior during reading. Most notably, the targeting mechanism based on the activation field can explain simple forward saccades, refixations, skipping, as well as small and large regressive saccades, without additional cognitive assumptions. Since foveal activation inhibits the independent random timer, foveal difficulty can inherently affect the current, as well as the upcoming fixation duration.

In summary, a variety of computational models of eye movements during reading exists. The models can generate complex fixation sequences on sentences, where fixation durations and fixation probabilities depend on properties of the given text. There are major divides concerning the order of word processing as well as the relationship between word processing and oculomotor control.

1.3 Parameter estimation

Properties of data generated by computational models greatly depend on the choice of parameter values. Yet, parameters are often arbitrarily chosen by hand according to theoretical constraints or rules of thumb[3], although—especially in complex models—small parameter variations bear the potential to yield substantial and unexpected effects (Engbert, Nuthmann, Richter, & Kliegl, 2005). Often few parameter configurations are explored in simulation studies, although reading data contains considerable interindividual variabil-

[3] As evident in formulations like "[..]we had assumed unrealistically long motor programming times in earlier versions of the model [..] However, we realized we had overdone it [..]" in a simulation study on preview benefits with the E-Z Reader model (Pollatsek, Reichle, & Rayner, 2006).

ity (Rayner, 1998). Individual scores in external tests of reading ability can explain more variation in fixation durations, especially in SFD and FFD, than general, established predictors like word length and word frequency (Everatt & Underwood, 1994; Kuperman & Van Dyke, 2011). However, when computational models are used to simulate experimental findings, interindividual variability is usually not considered.

To obtain parameter values that produce simulation results in agreement with experimental data, in the simplest cases researchers initially often turn to a grid based approach (e.g., Reichle, Rayner, & Pollatsek, 2003). Here a measure of the goodness of fit is calculated for set of equidistant points placed in the parameter space, allowing for easy comparison, visualization and interpolation. However, the approach quickly becomes infeasible with increasing dimensionality (i.e., number of free parameters), as the number of necessary evaluations grows exponentially. In consequence, parameters must be estimated algorithmically. Before discussing different methods, it is useful to specify what is meant by *goodness of fit*.

1.3.1 Goodness of Fit

Data coming from generative computational models often lends itself to the same means of analysis used with experimental data. One strategy is therefore to maximize the similarity of selected statistics between simulated and experimentally collected data (e.g., as in Engbert et al., 2002; Reichle et al., 1998), for example by least-squares estimation (Myung, 2003; Palestro, Sederberg, Osth, Van Zandt, & Turner, 2018). However, this often requires the simulation of large data sets which is costly in terms of computational power, and therefore computation time. Furthermore, different statistics likely evaluate different aspects of the model (Schütt et al., 2017), and combining several statistics usually involves the use of an arbitrary weighting function (but see Wood, 2010).

The alternative is to use a likelihood-based approach, which evades the pitfalls immanent in the usage of summary statistics by circumventing the step of data simulation. In turn it requires that the likelihood-function $L_M(\theta|y)$ of the generative model M is known or can be constructed.

Simply put, the likelihood quantifies the plausibility of observations for different parameter values as a function of the probability of making an observation under specific parameters:

$$L_M(\theta|y) = P_M(y|\theta). \tag{1.1}$$

While $P_M(y|\theta)$ describes the distribution of data y, given specific parameters θ for the model M, conversely, the likelihood $L_M(\theta|y)$ describes the distribution of parameters, given specific data. Using the likelihood in order to quantify the goodness of fit eliminates biases introduced by focusing on arbitrary statistics. This implies, that whenever a model can generate the observed data with nonzero probability, it is also possible to calculate a likelihood, making the likelihood approach viable for a broad class of models (Schütt et al., 2017).

1.3.2 Likelihood maximization

In maximum likelihood estimation (MLE) the aim is to find a set of parameters that maximizes the likelihood $L(\theta|y)$ for a data set y (Myung, 2003). When the likelihood function can be differentiated, this can be achieved analytically by setting the derivative to zero. However, this approach often is not possible, if the model involves many parameters or the likelihood function is highly non-linear. In such cases, many different optimization algorithms exist, which systematically search the parameter space for an optimal position. Typically they apply small changes to an arbitrary starting position repeatedly, and close in on a maximum over iterations, sometimes involving trial and error. This approach allows the exploration of large parameter spaces, which would be impossible to exhaust using, grid searches or pure Monte Carlo sampling.

A common problem in parameter estimation is posed by local maxima, which are spatially separated areas of higher likelihood within the parameter space, with areas of lower likelihood in between. Iterative optimization algorithms can become trapped in such areas. Although many algorithms can employ different ways in order to circumvent getting stuck in local maxima, the means of handling this problem remain heuristic and there is no perfect solution.

1.3.3 Bayesian inference

Instead of finding a single value or position in the parameter space, in the Bayesian approach a posterior distribution $P(\theta|y)$ over the parameters θ after the observation of the data y is calculated. The likelihood $L(\theta|y)$ represents constraints from the experimental data and the model, whereas a prior probability $Q(\theta)$ must also be specified, reflecting a-priori knowledge or beliefs on the model parameters. The choice of a prior distribution offers a possibility to incorporate auxiliary assumptions into the calculation, e.g., posterior distributions gained in a previous experiment or through a meta-analysis. However, when no prior information is available, broad and uninformative distributions should be used, to avoid biasing the results (Schütt et al., 2017). The posterior distribution is then given by

$$P(\boldsymbol{\theta}|y) = \frac{Q(\boldsymbol{\theta})L(\boldsymbol{\theta}|y)}{P(y)} \ . \tag{1.2}$$

To avoid having to calculate the denominator $P(y)$ one can use Markov Chain Monte Carlo (MCMC) methods (Gilks, Richardson, & Spiegelhalter, 1995), in turn yielding

$$P(\boldsymbol{\theta}|y) \propto Q(\boldsymbol{\theta})L(\boldsymbol{\theta}|y) \ . \tag{1.3}$$

MCMC methods use samples from the posterior based on local evaluations of likelihood and prior (Schütt et al., 2017). One such method is the random walk Metropolis-Hastings

MCMC algorithm (Hastings, 1970). Starting at an arbitrary position X_0 in the parameter space, a *proposal* Y_1 is calculated by updating the position with a random value drawn from some (e.g., Gaussian) distribution with $\mu = 0$ and some σ:

$$Y_1 = X_0 + \mathcal{N}(0, \sigma) \ . \tag{1.4}$$

In a second step, the posterior probability is calculated for the position and the proposal according to eq.1.3 to determine an acceptance probability:

$$\alpha_1 := \alpha(X_0, Y_1) := min\{1, \frac{Q(Y_1)L(Y_1|y)}{Q(X_0)L(X_0|y)}\} \ . \tag{1.5}$$

When the posterior probability of the proposal is greater the the posterior probability of the previous position, the proposal is accepted into the chain: $X_1 = Y_1$. If it is not greater, the proposal can still be accepted with the probability given by α. Otherwise the previous position is re-sampled, so that it is now contained twice in the chain: $X_1 = X_0$. After the decision on what value will be added to the chain, the process of generating a proposal based on the last chain value is simply repeated over and over. This simple random walk algorithm has interesting properties, namely that, given enough iterations, the distribution of collected samples will eventually converge to the posterior distribution. While it is not very efficient, many variations exist that improve performance (e.g., Laloy & Vrugt, 2012; Vihola, 2012).

1.4 The present studies

In the present studies the fields of experimental research, computational modeling and parameter estimation are combined. We first investigate parafoveal processing of lexical properties in an experiment with a modified $n+1$ boundary paradigm. Then, in an intermediate step, we develop the likelihood function of the SWIFT model and demonstrate the feasibility of parameter estimation based on simulated and experimentally recorded data in a Bayesian framework. Finally, we explore whether different mechanisms in the SWIFT model explain the effects found in the experiment, using simulations, based on parameter estimations on the participant level.

Although previous studies could show effects of parafoveal word processing on fixation durations (Risse, Hohenstein, Kliegl, & Engbert, 2014; Risse & Kliegl, 2014; Schotter & Leinenger, 2016), their experimental designs did not allow for a clear statistical distinction with respect to the nature of these effects. Therefore the aim of the first article was to clarify the relationship between the effects by improving on the experimental paradigm, which allowed for an unambiguous distinction between the effect of preview difficulty and preview validity.

The second article has a more general aim and is a necessary precursor to further work. We demonstrate the feasibility of data assimilation in the context of the SWIFT model of eye movements during reading. Beginning with a detailed description of the generative model, we develop a likelihood function using a computational approach. We then demonstrate the process of parameter estimation for a small, but demonstrative set of parameters for simulated data, successfully recovering known parameters from the estimation, using a Metropolis type MCMC algorithm. In a final step we estimate parameters for data from the previous experiment and compare statistics for simulations based on the obtained parameters with the empirical data.

Having established a workflow for incorporating experimental data into the parametric outfit of the SWIFT model, we revisit the preview effects in the third article. We combine the approach of parameter estimation with the exploration of different model variants and compare how well they can capture the effect patterns in the experiment. We employ an improved algorithm for parameter estimation and reduce the amount of data in a proper cross validation approach.

1.4.1 Stable preview difficulty effects in reading with an improved variant of the boundary paradigm.

The notion that word processing influences fixation durations during sentence reading is generally accepted (Henderson & Ferreira, 1990; Salthouse & Ellis, 1980). As during sentence reading not all words receive a fixation, processing can not be constrained to fixated words, and must be extended to adjacent, parafoveal words as well (McConkie & Rayner, 1975). The boundary paradigm (Rayner, 1975a, ; see see Section 1.1.4) has been extensively used to study the influence of various properties of parafoveal words on fixation durations after the boundary, where the finding of reduced fixation durations for valid previews, as compared to invalid previews, is usually explained as “preview benefit” due to successful trans-saccadic integration.

A number of findings (i.e., PoF effects) challenge this interpretation which extends to lower level word properties but typically excludes lexical processing of previews. Specifically, previous studies on frequency effects for invalid previews (Risse & Kliegl, 2012, 2014) make a case for parafoveal processing up to a lexical level. This exceeds the notion of mere integration as the sole source for the observed differences in fixation durations.

In Chapter 2 we aim to reproduce these findings with an alternative experimental design within the framework of the boundary paradigm, which allows for a statistically thorough discrimination between effects of preview validity and preview difficulty. Participants were presented with previews of three different difficulties (i.e., high, medium and low frequency classes), whereas all targets had the same difficulty (medium frequency class). This design eliminates the problems present in earlier experiments, where the observed effect sizes could have been confounded with asymmetric preview validity effects.

The results confirm the existence of independent effects of preview validity and preview difficulty for fixation durations on the target word after the boundary. We discuss the results in the context of SAS and GD modeling paradigms, and address alternative explanations, previously posed by Schotter and Leinenger (2016), with an additional time-course analysis based on the distributions of fixation durations.

1.4.2 Bayesian parameter estimation for the SWIFT model of eye-movement control during reading

In Chapter 3 we focus on data assimilation for a dynamical cognitive model of oculomotor control during reading. The SWIFT model (Engbert et al., 2002) of eye movements during reading accomplishes word processing by using an asymmetric processing gradient, centered on the fixation position. Every word that falls within this gradient is processed simultaneously, albeit with a speed depending on its eccentricity with respect to the position of the fovea, as well as the reading direction (hence the asymmetry). Individual words serve as targets for saccadic eye movements and are selected based on the progress of their processing. The temporal aspects of eye movements are governed by a set of hierarchical, stochastic timers which implement two distinct stages of saccade programming. The timers are initiated by a rhythmic generator process that is dynamically inhibited by the processing of the currently foveated word.

To achieve an optimal fit between model and empirical data, we seek a framework that circumvents the use of arbitrary statistics based on the model output and instead directly evaluates the likelihood of data, given the model. In the construction of the likelihood function we combine probability density approximation and pseudo-marginal likelihood estimation. In order to demonstrate the applicability of the likelihood function, we show changes of the likelihood for single parameter variations for a small subset of four parameters in simulated data, as well as the ability to recover a set of several parameters using an adaptive Metropolis Hastings MCMC algorithm (Vihola, 2012, ;see 1.3.3).

Finally, we apply the technique to empirical data collected in the experiment from Chapter 2, to estimate parameters on a participant level. In a critical check of the goodness of fit we compare typical summary statistics between the empirical source data and data simulated by using point estimates of the posterior parameter distributions. Although the estimation involved only four parameters, the means of simulated first pass fixation durations and probabilities are in good agreement with their empirical counterparts on the participant level.

1.4.3 Predictive modeling of the influence of parafoveal information processing on eye guidance in reading

In Chapter 4 we explore whether different mechanisms of inhibition by parafoveal words and oculomotor constraints triggered by display changes can account for the effects of preview difficulty and preview validity from Chapter 2 in the framework of the SWIFT model. In the model, saccade timing is only indirectly connected to word processing via inhibition of an otherwise autonomous saccade timer. Originally, only foveal words can exert inhibition, proportional to the fixated words processing progress. We extend the model to also include inhibition by parafoveal words as proposed by Henderson and Ferreira (1990) and explore a further variant, where parafoveal inhibition is delayed. Additionally, the model is provided a mechanism of saccade cancellation after a display change. In a further variation the saccade cancellation is coupled to preview processing, allowing for cancellation only while the preview is still in the first stage of processing at the time of the display change.

Using the empirical data gathered in Chapter 2, and the parameter estimation techniques developed in Chapter 3, we now estimate a large number of free parameters for the original model and the two model variants with parafoveal inhibition. Parameters are estimated on the participant level, now only using data from the control condition of the experiment in Chapter 2 (i.e., valid, medium frequency previews). In the next step, the resulting posterior distributions are used as a sampling reservoir in simulations of data for the three preview conditions in all model variants. We combine the variants of parafoveal inhibition with the variants of saccade cancellation, resulting in 9 different model configurations. For every combination the preview effects on the target word are calculated and compared to the experimental results.

Effects of preview difficulty emerged from the implementation of parafoveal inhibition and slightly improved, when inhibition was additionally delayed by 100ms. While preview validity effects were slightly present as a result of how the display change was implemented, they drastically increased in size, when saccade cancellation was added. Tying saccade cancellation to the first stage of preview processing reduced the effect size to a reasonable level. Taken together, the models with processing dependent saccade cancellation and parafoveal inhibition (delayed or immediate) showed the best agreement to the empirical data.

In conclusion we show that the SWIFT model of parallel, gradient based word processing can account for effects of lexical difficulty on the target word in a gaze contingent boundary paradigm, when the inhibitory influence of word processing is extended to include the upcoming parafoveal word. Simultaneously we demonstrate the capabilities of the Bayesian approach to likelihood based parameter estimation for small data sets and many parameters.

Chapter 2

Stable preview difficulty effects in reading with an improved variant of the boundary paradigm.

Sarah Risse and Stefan Seelig

Universität Potsdam, Germany

Running Head: Stable preview difficulty effects

published 2019 in *Quarterly Journal of Experimental Psychology, 72*(7), 1632–1645.
doi: 10.1177/1747021818819990

Abstract

Using gaze-contingent display changes in the boundary paradigm during sentence reading, it has recently been shown that parafoveal word-processing difficulties affect fixations on words to the right of the boundary. Current interpretations of this post-boundary preview difficulty effect range from delayed parafoveal-on-foveal effects in parallel word-processing models to forced fixations in serial word-processing models. However, these findings are based on an experimental design that, while allowing to isolate preview difficulty effects, might have established a bias with respect to asymmetries in parafoveal preview benefit for high-frequent and low-frequent target words. Here, we present a revision of this paradigm varying the preview's lexical frequency and keeping the target word constant. We found substantial effects of the preview difficulty in fixation durations after the boundary confirming that preview processing affects the oculomotor decisions not only via trans-saccadic integration of preview and target word information. An additional time-course analysis showed that the preview difficulty effect was significant across the full fixation duration distribution on the target word without any evidence on the pre-target word before the boundary. We discuss implications of the accumulating evidence of post-boundary preview difficulty effects for models of eye movement control during reading.

2.1 Introduction

What influences the decision of when and where to move the eyes during reading has long been of interest to psychologists and spawned a diverse field of research. In an influential study, McConkie and Rayner (1975) devised an experimental paradigm where, using an eyetracker, they manipulated the amount of visible text around the centre of fixation while the participants were reading normal sentences. By systematically varying the size of this moving window and by comparing measures like fixation durations and saccade distances to unhindered reading, it was found that the area of visible text necessary for normal reading, called the perceptual span, extends about four letters to the left and as much as 14 to 15 letters to the right of fixation (McConkie & Rayner, 1975; Rayner, Well, & Pollatsek, 1980) indicating the use of both foveal and parafoveal word information during reading.

Psychological models trying to explain the control of our eye movements in a task such as reading differ in their assumption about how word processing in the perceptual span is scheduled and whether attention is gradually distributed across the entire span or sequentially shifted from one word to the next during a fixation. While graded distributed attention (GDA) models predict parallel (lexical) processing of foveal and parafoveal words, sequential attention shift (SAS) models assume covert attention shifts to the parafoveal word only after lexical processing of the foveal word has finished. As a consequence, the models differ in their dynamics of processing foveal and parafoveal words inside the perceptual span and should, in principle, show at least subtle differences in how parafoveal processing affects the reader's eye movements.

A lot of research has focused on the type of information that can be extracted from parafoveal vision (see Hyönä, 2011; Schotter et al., 2012, for comprehensive reviews). However, little is known about the time course of information integration in the perceptual span, which would be most informative with respect to modelling eye movement control. The present study tests parafoveal processing effects across a target region spanning more than one word, and we discuss the compatibility of the results with current mathematical implementations of reading models based on SAS and GDA assumptions.

2.1.1 Parafoveal processing in the perceptual span

Using the gaze-contingent boundary technique (Rayner, 1975b), much has been learned about the impact of parafoveal word previews on reading. In such experiments, participants typically read sentences in which a predefined target word is initially masked with a different word or nonword. As soon as the eyes cross an invisible boundary, usually placed after the last letter of the word before the target (i.e., after the pretarget word n directly to the left of the target word $n + 1$), a display change occurs during which the target word is restored. This condition with invalid preview is then compared with the condition with valid preview in which the preview and target words are the same. As the

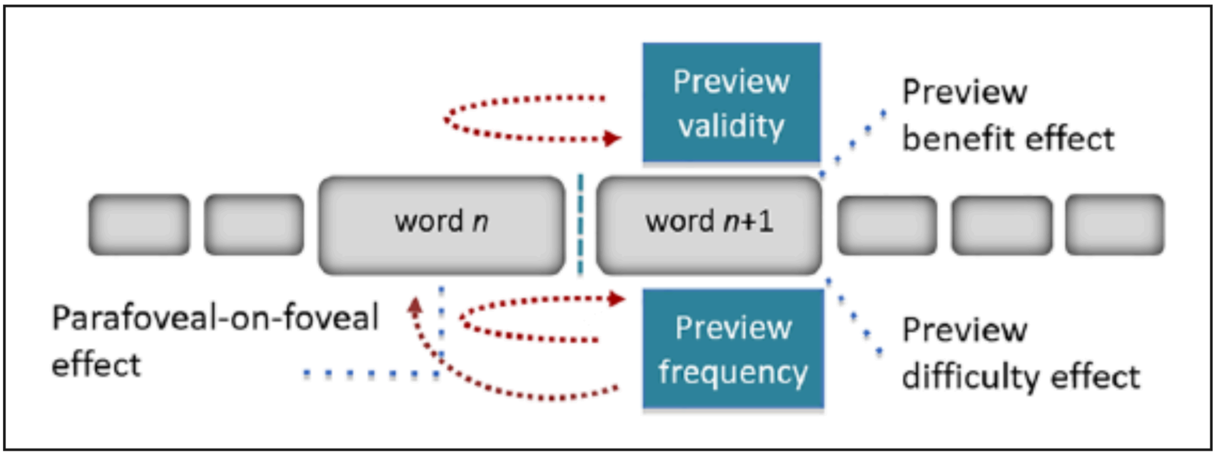

Figure 2.1: Illustration of parafoveal preview effects in the boundary paradigm manipulating preview of word $n+1$. The blue boxes contain the feature of the preview that has been manipulated. The red arrows reflect the spatio-temporal route via which the preview feature is assumed to affect fixation durations. Blue dotted lines indicate on which words the respective preview effects are measured. Varying the preview validity (valid vs. invalid) typically leads to preview frequency (high vs. low frequency) may result in parafoveal-on-foveal effects on word n and/or preview difficulty effects on word $n+1$. The dashed line between word n and word $n+1$ shows the location of the invisible boundary.

boundary is mostly crossed during a saccade, the display change should go unnoticed due to saccadic suppression (Matin, 1974; see also Slattery, Angele, and Rayner, 2011) and the previews are never subject to foveal inspection.

With this experimental technique, parafoveal preview of the target word has been shown to have a substantial effect on eye movements during reading (e.g., shorter fixations after valid as compared with invalid preview; see below for more details). Yet, preview effects are reported mainly for fixations after the boundary on the target word that was subject to the preview manipulation. Only under certain circumstances, they seem to also occur nonlocally on words before and after the target word. The precise time course of parafoveal information uptake and its integration with ongoing foveal processing is still unclear, and the mechanisms causing the observed local and nonlocal preview effects are disputed. In the following, we will briefly describe the empirical evidence for the different preview effects that have been obtained within boundary experiments (see 2.1 for an illustration), and then turn to their theoretical implications with respect to the debate about attention and the time course of foveal and parafoveal lexical word processing during reading.

Parafoveeal preview benefit. The best-established preview effect in the boundary paradigm is the parafoveal preview benefit that is observed in fixations on the target word after the eyes have crossed the boundary (see 2.1). Preview benefit is typically obtained when manipulating the validity of the preview that is the preview-to-target overlap of information. Readers spend significantly less time fixating the target word when the parafoveal preview shared information (e.g., the initial three letters) with the target word as compared with when it did not (Rayner et al., 1982). Preview benefit amounts to 30 to 50ms (Hyönä, 2011; Schotter et al., 2012; Vasilev & Angele, 2017) and belongs to the most reliable effects in eyetracking research on reading. It has been demonstrated for orthographic (Balota et al., 1985; Snell et al., 2017), phonological (Pollatsek et al., 1992), and even semantic (Hohenstein & Kliegl, 2014; Schotter, 2013) preview-to-target information overlap. Preview benefit indicates parafoveal word processing during reading and suggests that parafoveal information is integrated across saccades (Inhoff & Tousman, 1990). This trans-saccadic integration speeds up the processing of the foveated word, resulting in shorter fixation durations on the target word in case of valid preview (more information that can be integrated) as compared with invalid preview (less information that can be integrated). However, as integration can start only after the preview has been replaced with the target word, the preview benefit constitutes a rather late effect of parafoveal processing. More importantly, from the perspective of trans-saccadic integration, preview processing affects oculomotor control only indirectly. The temporal oculomotor decisions are solely based on the processing demand of the currently fixated foveal target word. However, the target word's processing demand is reduced by the amount of preview information that could be successfully integrated. Preview benefit can thus be viewed as a preview-reduced effect of the foveal word difficulty.

Parafoveal-on-foveal effects. In addition, preview processing may also have an effect on oculomotor control on its own. In boundary experiments, while fixating on the pretarget word n before the boundary, the parafoveal word $n+1$ to the right differs as a function of the preview condition. In the valid preview condition, the preview is the target word. But in the invalid preview condition, the target word is masked, often using a random-letter nonword, which should deviate substantially in processing difficulty from the valid preview condition. If the parafoveal preview is processed during pretarget fixations, difficult invalid previews may draw on more attentional resources than easier valid previews of the target word and postpone the decision of when to move the eyes away from the pretarget word.

Effects of the preview difficulty in pretarget fixations (i.e., on word n) are called parafoveal-on-foveal effects (see 2.1). They have been consistently reported for orthographic familiarity, and even for lexical frequency, in statistical analyses of eye movements while reading large text corpora (Kennedy & Pynte, 2005; Kliegl et al., 2006). In experimental designs, however, the results have been less consistent. Orthographic parafoveal-on-foveal effects have been shown for orthographic irregularities in parafoveal vision (e.g., words vs. nonwords; Inhoff et al., 2000) and higher-frequency orthographic neighbours (e.g., blue vs. blur; Snell et al., 2017). Lexical parafoveal-on-foveal effects, if at all significant, were very small and even inconsistent regarding their directions, sometimes showing increased and sometimes decreased pretarget fixations with difficult low-frequency (LF) previews (Hyönä and Bertram, 2004; but see Kennedy et al., 2002). A recent meta-analysis of 28 preview experiments reported strong evidence for the absence of lexical parafoveal-on-foveal effects of word $n+1$ (Brothers et al., 2017). Thus, whether the lexical processing difficulty of the preview modulates preboundary fixations is generally questioned.

Post-boundary preview difficulty effects. It has been shown that, under certain conditions, processing effects may lag behind and manifest during later fixations. For example, the lexical frequency of a currently fixated word n can affect the duration of subsequent fixations on the upcoming word $n+1$ (i.e., spillover effects, Rayner and Duffy, 1986; lag effects, Kliegl et al., 2006; foveal-on-parafoveal effects, Schroyens, Vitu, Brysbaert, and d'Ydewalle, 1999). Similarly, processing difficulties of the parafoveal preview that are expected to modulate fixations on the pretarget word n (i.e., showing parafoveal-on-foveal effects) may lag behind and affect the next fixation when the eyes have already crossed the boundary. As a consequence, parafoveal-on-foveal effects may be small and comparably weak on word n, as they only affect a small portion of the preboundary fixations. However, there might be stronger and more reliable preview difficulty effects in post-boundary fixations on word $n+1$ (see 2.1).

Risse and Kliegl (2012; see also Kliegl, Risse, and Laubrock, 2007) have first reported such effects investigating preview of word $n+2$, the word two words to the right of the boundary. Effects of the preview difficulty of word $n+2$ were found in fixation durations after the boundary on word $n+1$. Following up on this, they have demonstrated post-boundary preview difficulty effects also for parafoveal preview of word $n+1$ (Risse & Kliegl, 2014). In their boundary experiments, they combined high-frequency (HF; easy) previews with LF (difficult) targets and vice versa to test invalid preview conditions. In valid preview conditions, preview and target words remained the same (i.e., either the same HF or LF word). While they did not find any significant parafoveal-on-foveal effect of the preview difficulty in fixation durations on word n prior to the boundary, they found, in all experiments, a significant effect of the preview difficulty in fixation durations on word $n+1$ after the boundary. Difficult (LF) previews yielded an increase in fixation durations on the post-boundary word $n+1$, as compared with easy (HF) previews. As preview difficulty typically measures parafoveal-on-foveal effects on word n, finding the effect on word $n+1$ instead was interpreted as a *delayed parafoveal-on-foveal effect.*

2.1.2 Theoretical relevance of parafoveal-on-foveal effects

One reason for the ongoing interest in the controversy about parafoveal-on-foveal effects is that they are commonly taken as evidence for cross-talk in that "properties of a word present in the parafovea (i.e., not directly inspected), may influence the way a currently fixated word is processed" (Kennedy, 2008, p. 1). Given this definition, their existence would support parallel word processing, which is associated with models assuming that attention is gradually distributed across all words inside the perceptual span (i.e., GDA models). The statistical absence of parafoveal-on-foveal effects, to the contrary, is taken as support for models that postulate serial word-by-word processing in the perceptual span (i.e., SAS models; see Drieghe, 2011, for a detailed summary of this discussion).

At the same time, all contemporary models assume—at least to some degree—parallel processing of multiple words (e.g., by subdividing word processing into several hierarchical/serial stages, where early, low-level stages can already be concerned with the next word) and are, in principle, compatible with orthographic parafoveal-on-foveal effects. The theoretically interesting effects are thus lexical parafoveal-on-foveal effects that are typically investigated varying the frequency of the $n + 1$ preview. The above reported findings of preview difficulty effects confirm that parafoveal previews can be processed up to their lexical level during pretarget fixations. Moreover, this affects oculomotor control not only at the level of target selection (i.e., HF previews are skipped more often than LF previews; see Brysbaert, Drieghe, and Vitu, 2005) but also at the level of timing the next saccade (i.e., fixation durations are longer with LF previews than with HF previews). Yet, the effect on fixation durations seems to appear not on word n but is, relative to preview uptake (i.e., before the eyes crossed the boundary), delayed into later fixations (i.e., after the eyes crossed the boundary). To the extent that this finding results from cross-talk between processing the preboundary word n and the $n + 1$ preview in parallel, the absence of lexical parafoveal-on-foveal effects on word n as reported in Brothers et al. (2017) would not necessarily conflict with GDA model assumptions. In fact, the parafoveal-on-foveal effect on word n would just be shifted to word $n + 1$ (i.e., resulting in a delayed parafoveal-on-foveal effect). One plausible reason for such a position shift in observing the effect could be that the decreasing visual acuity in parafoveal compared with foveal vision delays lexical access of the preview such that its processing difficulty affects the oculomotor system later relative to foveal processing (see Lee, Legge, and Ortiz, 2003; Schiepers, 1980).

At the same time, the post-boundary preview difficulty effect can also be accounted for by SAS models. As Risse and Kliegl (2012, 2014) discussed, the neuronal delays of transmitting retinal information to higher cortical brain areas for lexical processing may lead to situations in which, during fixations on word $n + 1$, a saccade is programmed and executed still based on the old and meanwhile outdated information about the preview (see also Morrison, 1984). Thus, until the preview is updated with the target word, there is a small time window that allows preview frequency effects to occur on word $n+1$. This interpretation is fully compatible with the assumption of serial word processing during reading because lexical processing of word $n+1$ would have started during fixations on the pretarget word n, but only after lexical processing of word n would have been terminated, thereby avoiding any cross-talk between foveal and parafoveal processing.

2.1.3 An alternative interpretation in terms of forced fixations

Recently, Schotter and Leinenger (2016) proposed an alternative explanation in SAS models based on forced fixations. Basically, this extends the eye-brain-lag explanation described above to saccade programmes that were already started during fixating word n. Upon termination of processing word n, attention is shifted to word $n+1$ and processing of the preview can start while a saccade towards word $n+1$ is programmed. Given an easy word $n+1$ preview, the authors assume that parafoveal processing can occasionally proceed up to a level at which a second saccade programme to the next word $n+2$ is initiated. In most of these cases, this leads to the cancellation of the ongoing saccade programme and word $n+1$ will be skipped with the next eye movement. However, in some cases, the ongoing saccade programme to word $n+1$ will be in a stage at which it cannot be cancelled anymore and a fixation on word $n+1$ is forced. As a second parallel saccade programme to word $n+2$ is already in progress, such forced fixations on word $n+1$ will be short, and because they are more likely in case of easy HF previews and less likely in the presence of difficult LF previews, they have the potential to account for the observed preview difficulty effects on word $n+1$.

The latter interpretation in terms of forced fixations is compelling and points towards an interesting complexity within SAS models. Nevertheless, the current empirical evidence is likewise accounted for by GDA models. However, and more important for the present study, the critical evidence has been obtained in a paradigm that, to isolate effects of preview difficulty, introduced a potential bias as a consequence of asymmetries in preview benefit associated with different target word frequencies. As mentioned above, Risse and Kliegl (2012, 2014; see also Niefind and Dimigen, 2016; Schotter and Leinenger, 2016, Experiment 1) combined LF previews with HF target words and vice versa (see 2.2a). The orthogonal preview difficulty-by-validity manipulation allowed a valid assessment of both main effects and their interaction as long as the variables are independent of each other. However, as Inhoff and Rayner (1986) first showed, more preview is gained from an HF, than from an LF target word. As the preview difficulty main effect in the critical experiments was estimated by averaging the data from the two validity conditions, such asymmetric preview benefit could have resulted in a preview difficulty main effect mainly upon the selective preview benefit for HF target words (see 2.2c). In other words, the main effect of preview difficulty may have in fact resulted from statistical averaging across asymmetric validity levels.

Schotter and Leinenger (2016) tried to circumvent this issue reporting statistical tests for a nested design instead of the crossed 2×2 design. The resulting comparison was between HF and LF previews keeping the target word frequency constant. While keeping the target word frequency constant is the right way to go to achieve an independent test of the preview frequency, their experiments (except Experiment 2) still confounded the preview benefit: One level of the nested factor (e.g., LF preview/LF target) was always with valid preview whereas the other one was with invalid preview (i.e., HF preview/LF target). Moreover, while nesting is undoubtedly the better test of data with dependencies, simulations have shown that optimising designs in terms of collecting independent observations are always superior to nested designs with respect to type I error rate and

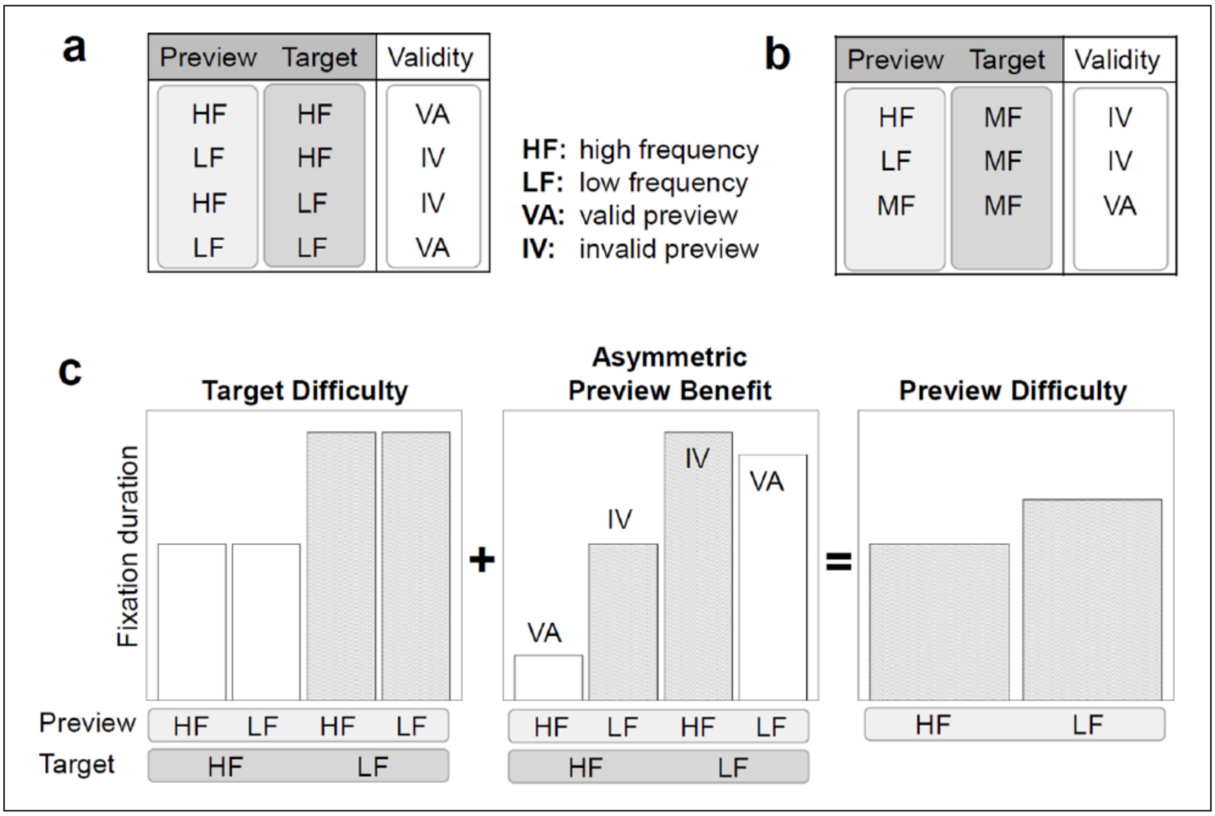

Figure 2.2: (a) Preview-to-target mapping in the experiments reported by Risse and Kliegl (2012, 2014), (b) preview-to-target mapping in the present study, and (c) illustration of the preview difficulty effect (right panel) as a by-product of target word difficulty (left panel) and potential preview benefit asymmetries (middle panel).

statistical power (Aarts, Verhage, Veenvliet, Dolan, & Van Der Sluis, 2014). Therefore, we revised the paradigm and proposed a design that tests the effect of preview frequency keeping the target word frequency constant [1].

2.1.4 The present experiment

In the present study, we used a modified experimental design that allowed a clear, non-confounded test of preview difficulty effects in fixation durations on the target word $n+1$ (see 2.2b). In a gaze-contingent display-change experiment, participants read single sentences while their eye movements were recorded. The sentences contained an invisible boundary before a target word, which was presented as one of three previews during fixations to the left of the boundary. Instead of combining easy (HF) previews with difficult (LF) target words and vice versa (plus the two valid preview conditions), we combined three levels of preview difficulty with only one level of target difficulty. Therefore, we systematically manipulated the frequency of the parafoveal preview (i.e., HF or LF), while

[1]The idea is similar to Experiment 2 in Schotter and Leinenger (2016) with the difference that they used nonword targets that were fixated after the boundary. This might have led participants to use strategies in their experiment that differed from normal reading.

keeping the frequency of the target word constant (i.e., a medium-frequency [MF] word). In addition to the two invalid preview conditions, the control condition tested the MF word as a valid preview for the MF target.

Given that the target word after the boundary was the same in all conditions, preview difficulty effects could not stem from interactions with differences in target word processing (e.g., asymmetric preview benefit). As a consequence, preview difficulty effects, if obtained in the present experiment, must be attributed to oculomotor decisions based on the processing difficulty in parafoveal vision before the eyes crossed the boundary. In line with previous research, we expected to find reliable parafoveal preview benefit in fixations on the target word. Target word fixations should be shorter in case of valid preview (i.e., MF preview with MF target) as compared with sentences with invalid preview (i.e., HF or LF preview with MF target). More importantly, we expected to find additional preview difficulty effects in the experiment with shorter fixations in case of HF as compared with LF previews. Finding such effects in fixations on the pretarget word n would make a strong case for lexical parafoveal-on-foveal cross-talk. However, with respect to previous research, we expected preview difficulty effects mainly on the target word $n + 1$ after crossing the boundary. To decide whether such effects reflect delayed effects of parallel processing (e.g., delayed parafoveal-on-foveal effects) or forced fixations during serial processing is beyond the scope of the present study. Here, the goal is to provide a conservative test of the critical effect with an optimised methodology on which future research can build upon.

2.2 Method

2.2.1 Subjects

Thirty-four young adults (gender: $n_{female} = 20$, $n_{male} = 14$; age: $M = 21$years, $SD = 4$) were tested in a 45-min session. Participants were university or high school students from the Potsdam community and were compensated with course credit or 7 € for their attendance. All participants had normal or corrected-to-normal vision, which was assessed with the *Freiburg Visual Acuity Test* (Bach, 2007), and signed informed consent before beginning the experiment.

2.2.2 Sentence Material

The experiment consisted of 114 simple structured sentences with six to 12 words and an average length of nine words ($SD = 1$). A two-word target region was embedded in each sentence. The preboundary word n was an adjective, which ranged from four to 11 letters ($M = 6$, $SD = 2$), and occurred at positions two to eight in the sentence ($M = 4$, $SD = 2$). Word n was followed by the target word $n + 1$, a noun that ranged from five to seven letters. Forty sentences contained a five-letter noun, 38 sentences a six-letter noun,

Table 2.1: Summary statistics for preview/target word $n+1$ and pretarget word n. Frequency is defined as the occurrence of a word per million. Predictability is the probability of correctly guessing a word in a cloze task.

	Frequency word $n+1$			N letters		Predictability word $n+1$			Letter overlap (%)	
	HF	MF	LF	n	$n+1$	HF	MF	LF	HF-MF	LF-MF
M	107.7	29.1	2.4	6	6	.01	.04	0	30.0	29.2
SD	55.0	4.3	1.4	2	1	.04	.08	0	19.0	18.6
Minimum	50.2	20.6	.02	4	5	0	0	0	0	0
Maximum	362.1	40.0	5.0	11	7	.25	.5	0	85.7	83.3

HF: high frequency; MF: medium frequency; LF: low frequency; *M*: mean; *SD*: standard deviation.

and 36 sentences a seven-letter noun. Word $n+1$ was never the last word in a sentence and occurred at positions three to nine ($M = 5$, $SD = 2$). The critical manipulation concerned the frequency of the word $n+1$ preview. Each sentence frame enabled an HF (easy), LF (difficult), and an MF (medium) word at position $n+1$. Word lengths of the target and its previews were matched for each sentence and none of the previews resulted in a semantic violation. HF words had an average frequency of 108 per million ($SD = 55$), LF words averaged 2.4 per million ($SD = 1.4$), and MF words averaged to 29 per million ($SD = 4.3$). All word norms were retrieved from the dlexDB database based on the *Digitales Wörterbuch der Deutschen Sprache des 20. Jahrhunderts* corpus (Geyken, 2007; Heister et al., 2011). For a summary of characteristics of words in the target region, see Table 1.

In addition, cloze task predictability norms for each possible word $n+1$ (HF, LF, and MF words) were collected online. Fourteen participants, who did not take part in the reading study, read the 114 experimental sentences in randomised order up to word n and guessed the upcoming word $n+1$. As summarised in Table 1, MF words were correctly guessed in 4% of the sentences (averaged across sentences and subjects) and HF words in 1%. LF words were not correctly guessed once. The critical previews were thus extremely low predictable and the sentences can be considered neutral (cf. Schotter & Leinenger, 2016).

2.2.3 Apparatus and procedure

The sentences were presented horizontally on a 22-inch Iiyama Vision-Master Pro 514 monitor with a resolution of 1024 × 768 pixels and a monitor refresh rate of 150Hz. Using a chin-rest, the distance to the monitor was kept constant at 60 cm. Monocular right eye gaze position was recorded using the Eye-Link 1000 Tower-Mount-System (SR Research, Ontario, Canada) with a sampling rate of 1000Hz. Sentences were presented in black monospace font (Courier) with fontsize 14 (letter x-height: 9 pixels) on a white background with each letter spanning 0.34°of visual angle.

Participants were instructed to read for comprehension, which was assessed randomly after one third of the trials with a three-alternative multiple-choice question. Comprehension accuracy was high ($M = 98\%$, $SD = 3$). Each experiment started with the collection of the participant's informed consent. The participants then had to perform the Freiburg Visual Acuity Test (Bach, 2007) and determine their dominant eye. They were thereafter familiarised with the apparatus and procedure and provided with a written instruction. Calibration used a standard nine-point grid, and recalibrations were done every 15 trials and whenever necessary. Sentences were presented horizontally at the vertical midline of the monitor. The starting position of each sentence was fixed at 40pixels offset from the left monitor border. Each trial started with a fixation point on the left side of the horizontal midline that indicated the optimal viewing position (i.e., word centre) of the first word of each sentence. Drift correction was applied in the centre of the screen in case that gaze detection failed for 100ms. Recalibration followed after three successive failures. After a successful initial fixation, the sentence was displayed on the monitor. The presentation of the sentence was completed with the fixation of a dot in the lower right corner of the screen.

The experiment started with six practice sentences to familiarise the participants with the procedure, followed by 114 experimental sentences. To manipulate parafoveal preview, the boundary paradigm was implemented, inserting an invisible boundary at the end of the last letter of word n. The next word $n + 1$ changed contingent on the reader's eye position. Prior to the boundary, word $n + 1$ was either an HF, LF, or MF word. As the eye crossed the boundary, it was replaced with the MF word. Hence, each sentence was presented in one of three conditions: (1) HF–MF, (2) LF–MF, (3) MF–MF. In condition (3), preview and target words were identical, thus word $n + 1$ was replaced by itself and participants obtained valid preview. In conditions (1) and (2), preview and target words differed from each other. Thus, participants received invalid preview of the target word. Experimental conditions were counterbalanced across participants, and sentence presentation was randomised. Participants were asked to fill out a short questionnaire after the experiment, in which they had to state if they had noticed any display changes during the experiment. On average, self-reports of display changes amounted to 13% of sentences. The critical preview effects did not covary with the self-reported display-change awareness of the individual participants. Therefore, analyses were conducted based on data of all participants.

2.2.4 Data analysis

Saccades were detected offline in the eye movement records using the algorithm by Engbert and Mergenthaler (2006). Fixation durations (i.e., the inter-saccadic intervals) were then divided into first-pass single fixation durations (SFD; cases in which a word was only fixated once), first fixation durations (FFD; fixations associated with the first encounter of the word) and gaze durations (GZD: sum of all fixations on a word until first departure). If words were (re-)fixated after the gaze had already been to the right of the word, this was considered the second pass of reading. Second-pass fixation durations were added to the GZD constituting the total viewing time (TVT) on the word.

In all, 9% of the sentence data were discarded due to tracker loss and an additional 17% due to invalid display changes (i.e., target word replacement finished after fixation onset on the target). Individual fixation durations were considered outliers if they were shorter than 10ms or longer than 800 ms[2]. 2 Fixation duration measures that contained at least one outlier fixation were removed from the dataset (0.33% of word-based duration measures in the target region). This left 2.772 valid trials (including first- and second-pass reading) for analyses on word n and 2.783 trials for analyses on word $n + 1$.

Statistical analyses were performed with linear mixed models (LMMS; Baayen, Davidson, & Bates, 2008) using the *lmer* programme (lme4 package; Bates, Mächler, Bolker, & Walker, 2015) in the R environment (R Foundation of Statistical Computing R Core Team, 2013). Separate LMMs were estimated for each dependent variable (DV) controlling for random variances by submitting subjects and items as crossed random factors. In addition, and theoretically more important, the three preview conditions were compared using a priori Helmert contrasts. The preview difficulty contrast (c1) tested the HF preview condition (HF: –1) against the LF preview condition (LF: + 1). For the preview benefit effect, the preview validity contrast (c2) tested the average of the HF and LF preview conditions (HF: –1; LF: –1) against the MF preview condition (MF: +2). With these two fixed effects centred around zero, the intercept of the models reflected the grand mean of the DV across subjects, items, and conditions.

Fixation durations were log-transformed to increase accordance with the normality assumption. Note that the results did not differ when the models were estimated on the untransformed variables suggesting that the statistical results were stable and not dependent on parametric assumptions. For each fixed effect, t values were computed as the ratio of estimate (b) and standard error (SE) and are typically considered significant at $p < .05$, if $|t| > 1.96$. However, to be conservative (e.g., because we ran multiple analyses on several different fixation duration measures, see Von der Malsburg and Angele, 2017), t values should be substantially larger.

We were mainly interested in the theoretically motivated model described above. However, we ran additional models to check the reliability of the results in our main model. Mainly, we fitted the maximum models estimating random contrast slopes for both subjects and items (see Barr, Levy, Scheepers, & Tily, 2013, for more details). All models converged and showed very similar fixed effect parameters with no changes in significance

[2]We used a cutoff for short fixation outliers of 10ms instead of the typical 80ms to increase the chance of keeping brief fixations in the fixation distribution that might be the result of eye-brain lag (Morrison, 1984) or forced fixations (Schotter & Leinenger, 2016)

Table 2.2: Target word $n + 1$. Condition means(M) and standard deviations (SD) of various fixation measures (upper panel) and fixed effects (estimated slopes and t values) from linear mixed-effects models (lower panel). See text for more details.

Preview	FFD		SFD		GZD		TVT		SKP	
	M	SD	M	SD	M	SD	M	SD	M	SD
HF	251	89	254	88	272	107	338	153	.095	.29
LF	274	98	282	98	296	104	350	149	.052	.22
MF	233	74	236	73	246	85	287	137	.097	.30
	b	t	b	t	b	t	b	t	b	z
c1[pvD]	.04	**5.25**	.05	**7.14**	.05	**6.52**	.02	**3.02**	-.36	**-3.92**
c2[pvD]	-.04	**-7.62**	-.04	**-9.34**	-.04	**-10.4**	-.06	**-12.9**	.13	**2.61**

FFD: first fixation durations; SFD: single fixation durations; GZD: gaze duration; TVT: total viewing time; pvD: preview difficulty; pvV: preview validity; SKP: skipping probability.
The bold values indicate that t- (and z-)values are significant (with $p < .05$).

pattern. As we are not interested in individual differences in the present study, we report the results from the theoretical model only, $log(DV) \sim c1 + c2 + (1|subject) + (1|item)$.

2.3 Results and discussion

2.3.1 Main analysis: fixation durations on target word n+1

The results from the target word main analysis were unambiguous and clear-cut. Table 2.2 shows the LMM results for both first- and second-pass fixation duration measures and the associated condition means. Word $n + 1$ was fixated once during firstpass reading (i.e., in a single fixation) in 2,419 cases (87%) and was refixated in first-pass only in 255 cases (9%).[3] The remaining 4% were trials in which word $n + 1$ was skipped during first-pass reading but then revisited later during second-pass reading. In the following, we will mainly refer to the results from the analysis of FFD (which were mainly single fixations), as the first fixation after the boundary should be most critical with respect to delayed effects from the pretarget word. Note, however, that the effects replicated across the other fixation measures.

As Figure 3 (left panel) illustrates, the preview difficulty contrast was significant. Specifically, fixation durations were longer, if the parafoveal preview had been an LF as compared with an HF noun. Note that this preview difficulty effect was measured on

[3]The number of first-pass refixations on word $n+1$ were very low in this experiment so that measures such as the first fixation of multiple fixation cases or the refixation probability were not investigated with respect to effects of preview condition.

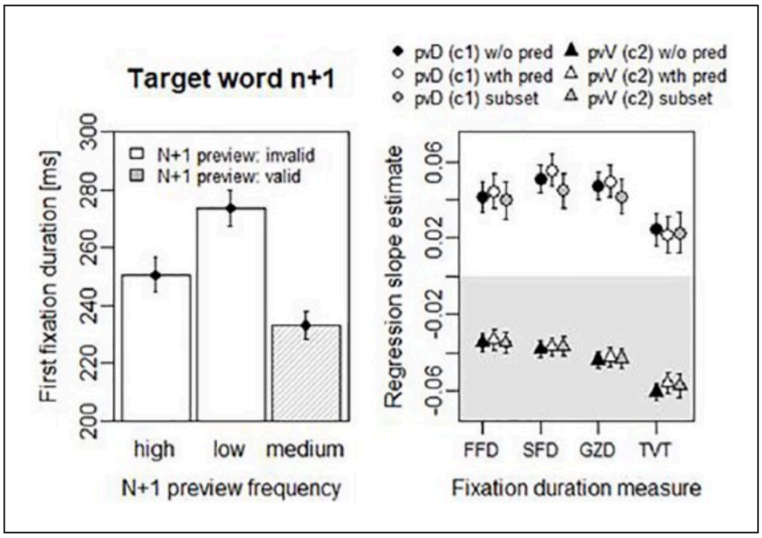

Figure 2.3: Left panel: Average first fixation duration on the target word $n + 1$ as a function of the preview difficulty. Error bars represent the 95% confidence intervals. Right panel: Estimated regression slopes and their standard errors for preview difficulty (pvD) and preview validity (pvV) effects of linear mixed-effects models with and without predictability as covariate and with a subset of data only including sentences in which all previews had a predictability of zero. Estimates are depicted for different fixation durations on word $n + 1$ (see text for more information).

the target word after the preview had been replaced and the preview was not visible in parafoveal, let alone foveal, vision. More importantly, as the preview was in both conditions changed into the same MF target word, this difference must reflect a pure effect of the processing difficulty of the preview before the boundary.

In addition, the preview validity contrast was significant as well. Both HF and LF preview conditions implied a display change and thus provided invalid preview for the target word, and their average fixation durations were longer than in the MF preview condition without display change (i.e., valid preview). Thus, the present experiment showed clear evidence for the existence of both preview benefit and preview difficulty effects in the same fixation duration measure on the same word after the boundary.

2.3.2 Control analyses: predictability and orthographic similarity covariates

Next to a word's lexical frequency, other linguistic variables such as the word's length or its predictability influence word recognition and many of them are correlated. While word length for previews and targets was matched and thus controlled in the present experiment, the predictability between HF, LF, and MF words might have differed. As predictability is the probability of guessing a word from the sentence context, differences

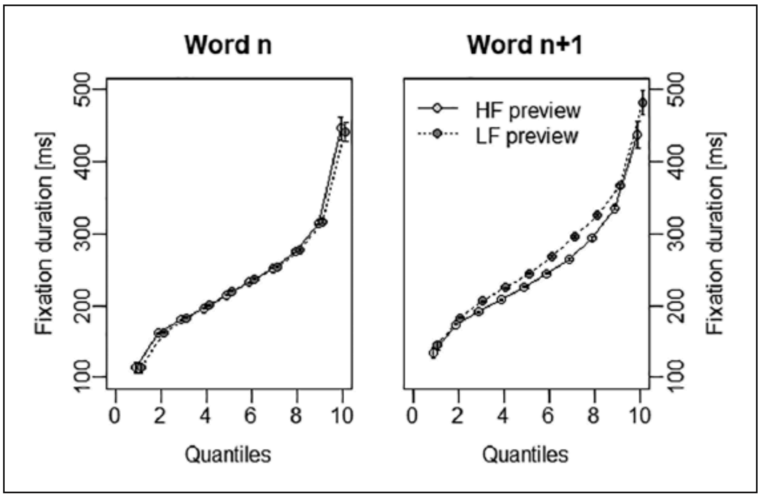

Figure 2.4: Quantile plots for last fixation on word n (left panel) and first fixation on word $n+1$ (right panel) as a function of high frequency (HF) and low frequency (LF) previews. Quantile means are plotted with a little offset in quantiles making the comparison between preview conditions easier for similar mean values. Fixation durations were binned for each preview condition pooled across participants. Error bars are very small and represent the 95% confidence intervals.

between previews may be accounted for solely by knowledge-based predictions without needing to assume any parafoveal processing of the preview (i.e., HF previews may be better predictable than LF previews resulting in shorter fixation durations).

Although the cloze predictabilities in the present study were very low, the 1% difference between HF and LF previews was significant, paired, one-sided t test: $t(113) = 3.29$, $p < .001$. We therefore ran additional LMMs including the cloze predictability of the $n+1$ previews as a covariate (logit-transformed and centred on zero). Estimating the effect of predictability alone (as main effect) and in combination with both preview contrasts (as interaction effects) left the model rank deficient because there was no variability in predictabilities for LF previews (all cloze predictabilities equal zero). Dropping the redundant interaction between predictability and the preview validity contrast, the predictability main effect reached significance in the TVT analysis ($b = -.02$, $t = -2.10$) and its interaction with the preview difficulty contrast was marginally significant in the analysis of SFD ($b = .03$, $t = 1.95$). In all other fixation analyses, the predictability had no reliable effect (all $|t| < 1.46$). The critical test, however, was to see whether the cloze predictability would explain the preview difficulty effects observed in the present study. Yet, submitting cloze predictability to the LMMs did not change either the preview difficulty contrast (c1) or the preview validity contrast (c2). Figure 2.3 (right panel) summarises the estimates and their standard errors for each contrast of the control analysis (white symbols) in comparison with the main analysis (black symbols). We also ran the main model on a subset of data excluding all sentences in which the predictability of one of the previews was larger than zero. The effect sizes of the preview difficulty and the preview validity contrast remained almost unaffected (see Figure 2.3, right panel, grey symbols).

Another variable that could possibly explain the effects of preview difficulty may be the orthographic similarity between preview and target word. It could have been the case that, by chance, HF previews shared more letters with the MF target words than LF previews. To test this, we computed the letter overlap between HF previews and MF targets as well as LF previews and MF targets. There was letter overlap of about 29% between previews and target words, but this did not differ between HF ($M = 30\%$, $SD = 19$) and LF previews ($M = 29\%$, $SD = 19$). More importantly, estimating an LMM on only the display-change conditions and with letter overlap as covariate did not explain significantly more variance than the baseline model without letter overlap and did not affect the estimates of the preview difficulty effect (see Table S1 in the online Supplemental Material). Thus, we consider the preview difficulty effect on word $n + 1$ a genuine effect that is not easily explained by the predictability of the preview or its orthographic similarity with the target word, but rather originates from processing difficulties associated with its lexical frequency.

2.3.3 Supplementary post hoc analyses on parafoveal-on-foveal effects

To test whether the significant preview difficulty effect on the target word $n+1$ could be interpreted as a delayed parafoveal-on-foveal effect spilling over from the pretarget word n as predicted from a strict parallel processing perspective, we conducted a supplementary analysis on the time course of this effect. Therefore, we selected a subset of fixation durations containing the last fixation on the pretarget word n and the first fixation on the target word $n + 1$ of each sentence, in which word n and word $n + 1$ were read in sequence. In case the preview difficulty effect on word $n + 1$ was a delayed effect from word n, longer pretarget fixations should increase the chance to find those effects already immediately on word n and not only delayed on word $n + 1$. Similarly, a delayed effect from the pretarget word should manifest already in short fixations on the target word.

We assessed the distribution of fixation durations before and after the boundary with quantile regressions using the quantreg-function (Koenker, 2015) in R (R Foundation of Statistical Computing, 2016). Figure 4 shows the fixation duration averages for each quantile (i.e., containing the first 10%, the second 10%, the third 10%, etc. of speed-ranked fixations), split by conditions of HF and LF previews, respectively. Quantile regressions showed nonsignificant preview difficulty effects on word n for all quantiles (all absolute t values < 1.42). However, they confirmed a significant preview difficulty effect on word $n + 1$ from the second until the last quantile (all ts > 2, all ps $< .04$). The preview validity effect on word $n+1$ was significant from the third until the last quantile (all absolute ts > 4, all ps < .001). See Tables S2 and S3 in the online Supplemental Material for more details on the quantile regressions.

In summary, contrary to predictions from a strict cross-talk assumption, fixation durations on the pretarget word seemed to not differ at all with respect to the frequency of the parafoveal preview (left panel), neither for short nor for long pretarget fixations. However, in fixation durations on the target word (right panel), the effect of the preview difficulty was evident already in the shortest fixations and was increasing across the entire fixation duration distribution. We will discuss potential implications in the "General discussion."

2.4 General discussion

The present study tested whether preview difficulty effects on word n+1 in gaze-contingent display-change experiments are genuine. Alternatively, they could have been artefacts from asymmetries in preview-to-target mappings that bias preview benefit and imitate a preview difficulty effect. Keeping the target word frequency constant instead of crossing its frequency with the frequency of the preview, as has been done in previous research, we replicated longer target word fixations after difficult LF previews as compared with easy HF previews. In line with that previous research, the preview difficulty affected fixation durations only on word $n + 1$ after the boundary and not on word n before the boundary. In the following, we will discuss how the results relate to current reading models, and what their temporal pattern suggests about the oculomotor control processes involved in reading.

2.4.1 Strong evidence for preview difficulty effects on word n+1

Reading sentences in which easy (HF) previews of word $n + 1$ were changed into difficult (LF) target words (or vice versa) when moving the eyes from word n to word $n+1$, several studies have now reported evidence for effects of the preview difficulty on oculomotor control (Niefind & Dimigen, 2016; Risse & Kliegl, 2012, 2014; Schotter & Leinenger, 2016). In those studies, fixation durations on word $n + 1$ were longer if its previous parafoveal preview had been a difficult (LF) word compared with an easy (HF) word, and this was independent of the present processing difficulty of the currently fixated word in foveal vision. Surprisingly, such preview difficulty effects were observed only on the target word $n+1$ after the boundary and not immediately on the pretarget word n before the boundary. Due to the theoretical relevance of preview difficulty effects that we will discuss in more detail below, it is important to make sure that such effects genuinely exist and are not an artefact of the method used to investigate them.

In fact, the change into a target word of opposite processing difficulty has at least two disadvantages. First, the later foveal processing difficulty may counteract the parafoveal processing difficulty and thus obscure preview difficulty effects even though they may exist. A difficult LF target word might place such a high demand on the processing system that any signal from the previously processed easy HF preview would basically be ignored. Second, averaging the preview difficulty across conditions in which preview and target word were also sometimes identical (i.e., pooling valid and invalid preview conditions of the same preview difficulty) could mimic preview difficulty effects although not existent. As Inhoff and Rayner (1986) showed, the benefit in fixation durations is larger for valid preview of HF compared with LF target words, and this asymmetry could have disproportionately reduced the fixation durations on word $n + 1$ in case of valid HF preview. Therefore, the main goal of the present study was to replicate preview difficulty effects in the boundary paradigm using an optimised method in which the target word frequency was kept constant.

The results were straightforward: We found reliable preview benefit on word $n + 1$ with shorter fixation durations after valid compared with invalid preview, confirming the most important benchmark result in boundary experiments. Parafoveal preview benefit amounted to $29ms$ in FFD, $32ms$ in SFD, $38ms$ in GZD, and $57ms$ in TVT. Although its size was in the lower range of what is typically reported (e.g., 30-50 ms; see Vasilev & Angele, 2016), we consider it reasonable as we dissociated a second reliable source of preview effects on word $n + 1$ that is in most classical boundary experiments confounded with the preview benefit: Fixation durations on word $n + 1$ were up to 29ms longer in SFD (23ms in FFD, 24ms in GZD, 11 ms in TVT) if the preview of the currently fixated word $n + 1$ was an LF word rather than an HF word. Thus, we replicated an effect of the preview difficulty on word $n + 1$ in an experiment in which it can be unambiguously attributed to a direct influence of the preview processing on oculomotor control. It is noteworthy that there was no such effect of the preview difficulty in fixating the pretarget word n. We will now turn to what this means with respect to the debate on attention, word recognition, and eye movement control during reading.

2.4.2 Serial versus parallel computational reading models

The two major theoretical accounts to explain eye movements during reading differ in that they either assume SAS accompanied by serial word processing or GDA allowing parallel word processing in the perceptual span. To test these very general theoretical assumptions against each other, one needs explicit models that implement these ideas and combine them with mechanisms that guide the oculomotor decisions of when and where to move the eyes. In case of the SAS approach, this has been successfully done in the E-Z Reader model (Reichle et al., 1998); computational models based on GDA are SWIFT (Engbert, Nuthmann, Richter, & Kliegl, 2005) and GLENMORE (Reilly & Radach, 2006).

All these models can account for preview benefit in fixations on the target word and do this in basically the same manner. They all allow for a certain amount of parafoveal preprocessing of the target word by either shifting or distributing attention to the word in the parafovea before the eyes will fixate it. A valid preview in parafoveal vision thus enables a head-start of processing the target word compared with any invalid preview and reduces the amount of target word processing that has to be accomplished on fixation. However, the models' mechanisms differ substantially when it comes to explaining parafoveal-on-foveal effects on word n or preview difficulty effects on word $n + 1$. Moreover, one needs to distinguish model mechanisms that are intended to account for certain effects from those that happen to do so because of interactions with other mechanisms in the model architecture. The latter are not less intriguing or of minor importance; however, they may not directly report on the plausibility of serial or parallel processing during reading.

Mechanisms to account for parafoveal-on-foveal effects on word n. Still today, the controversy about parafoveal-on-foveal effects is often summarised as follows: Parallel word-processing models assume cross-talk between foveal and parafoveal lexical processes and thus predict (lexical) parafoveal-on-foveal effects on the preboundary word n, whereas serial word-processing models do not (see Brothers et al., 2017; Drieghe, 2011). The increasing amount of null-findings with respect to such effects (see the Bayes Factor meta-analysis results by Brothers et al., 2017, but also the present results) is interpreted as evidence for SAS serial processing models and against GDA parallel processing models. However, this view widely ignores the complexity of the interactions between attention allocation and oculomotor control mechanisms in the models. For example, while serial one-word attention shifts in E-Z Reader prevent parafoveal processing to interfere with foveal word processing, imprecisions in the saccade system (cf. McConkie et al., 1988) can result in situations in which the parafoveal word is processed while the word to its left is erroneously fixated. Given certain word lengths, mislocated fixations can amount up to 30% in fixations measured on a word (Engbert & Nuthmann, 2008). Such mislocated fixations have been proposed to explain parafoveal-on-foveal effects in the E-Z Reader model (e.g., Drieghe et al., 2008; Rayner, Warren, Juhasz, & Liversedge, 2004). In fact, Reich and Cotter (2015) have shown that simulating systematic and random errors in fixation locations in the E-Z Reader model can lead to parafoveal-on-foveal effects. Thus, the interaction between SAS and oculomotor error calls into question the common claim that the strictly serial attention shifts in E-Z Reader predict absence of parafoveal-on-foveal effects.

At the same time, current versions of the SWIFT model, although employing parallel word processing across the perceptual span, do not result in cross-talk between foveal and parafoveal word processing. This is due to the fact that only the foveal processing difficulty affects the timing of the next saccade programme (i.e., via foveal inhibition). While parafoveal words are processed in parallel with the word in foveal vision, their difficulty affects the oculomotor system only via saccade-goal selection (for more details, see Risse & Kliegl, 2012, 2014). However, simulations with the SWIFT model of reading in the $n + 2$ boundary paradigm have shown that the effects on saccadic selection were not strong enough to produce significant parafoveal-on-foveal effects (Risse, Hohenstein, Kliegl, & Engbert, 2014). Thus, the current SWIFT implementation (or the particular set of best-fitting parameters) calls into question the common notion that GDA models always predict substantial parafoveal-on-foveal effects.[4]

Mechanisms to account for post-boundary preview difficulty effects. Risse and Kliegl (2012) have shown that preview difficulty effects on word $n + 1$ are unlikely a result of mislocated fixations. In experiments in which the frequency of the preview and the later target word of a word two words to the right of an invisible boundary (i.e., word $n + 2$) was crossed, they investigated fixation durations on the three-letter word $n + 1$ after the boundary and before the target word $n+2$. As word $n+1$ was short and easy to process, it was often skipped during reading the sentences (up to 50%), and fixations landing on it were likely mislocated and intended to land on word $n + 2$ instead. In SAS models, such mislocated fixations should go along with processing the intended fixation target, which would be word $n + 2$ in the present scenario. However, the results showed no influence of target word $n + 2$ processing in fixation durations on word $n + 1$ (i.e., no target frequency effects, no preview benefit) but only effects of the word $n + 2$ preview (i.e., preview difficulty effects). These results suggested that the duration of fixations on word $n + 1$ was not determined by processing word $n + 2$ while unintendedly fixating word $n + 1$ but was driven directly by the processing difficulty of the earlier preview. As preview difficulty effects before the boundary would have indicated parafoveal-on-foveal effects, finding them after the boundary led to calling them delayed parafoveal-on-foveal effects. However, if mislocated fixations do not account for this effect, how can a strictly serial model like E-Z Reader explain such a nonlocality?

Schotter and Leinenger (2016) proposed an explanation based on forced fixations in the E-Z Reader model: The interaction between covert serial attention shifts and a two-stage process of saccade programming to the next word in sequence may lead to situations in which a word that would in principle be skipped (i.e., not selected for being the next fixation target) is yet fixated. A forced fixation immediately triggers the start of a new progressive saccade programme away from the involuntarily visited word, already before the word is fixated. Thus, the duration of forced fixations are expected to be rather short and independent of any ongoing lexical word recognition processes. As on average, forced fixations should happen more frequently while processing HF previews than LF previews, they could explain the preview difficulty effect on word $n + 1$. From the durations of E-Z Reader's word processing and saccade programming stages, Schotter and Leinenger predicted that forced fixations should primarily be observed among the short fixations on the target word below 200ms. In contrast to that, the present results showed preview

[4]Note that other graded distributed attention models such as Glenmore make different predictions with respect to parafoveal-on-foveal effects as they realise different word processing (Reilly & Radach, 2006).

difficulty effects significant across the full fixation distribution (except the first quantile containing the shortest fixations). Whether this is solid evidence against the forced fixations account, however, can only be answered with model simulations showing that forced fixations in E-Z Reader fail to generate preview difficulty effects in the tail of the fixation duration distribution.

In a similar vein, simulations of preview difficulty effects in the SWIFT model would be required to make reliable assertions about the behaviour of this particular implementation of a parallel GDA model. The present absence of a link between parafoveal word processing and the timing of the next saccade programme (i.e., only foveal inhibition) renders such effects rather unlikely. In fact, SWIFT simulations of parafoveal processing in the $n+2$ boundary paradigm have shown that the preview difficulty effect on word $n+1$ was evident only when it was additionally assumed that a certain percentage of saccade programmes were cancelled when previews were replaced by the target word (Risse et al., 2014).[5] One reasonable way how to account for preview difficulty effects in fixations after the boundary could be the implementation of processing cross-talk into the SWIFT model by modelling foveal together with parafoveal inhibition. Besides the strongly diverging theoretical assumptions behind the two implementations, preview difficulty effects in SWIFT would differ from preview difficulty effects in E-Z Reader in that parafoveal inhibition would affect each fixation on word $n + 1$ to a certain degree whereas forced fixations refer to only a subpopulation of fixations on word $n + 1$. How this difference could be tested empirically remains an intriguing question for future research.

2.5 Conclusion

The present study confirmed the existence of preview difficulty effects in the boundary paradigm, excluding alternative explanations based on asymmetric preview benefit. Interestingly, the preview difficulty effect was obtained only on the target word $n+1$ with no evidence on the pretarget word n. In addition, the preview difficulty effect on word $n+1$ was significant across the entire fixation duration distribution, being reliable for short as well as for long fixations on the target word. While the spatial pattern of preview difficulty effects (i.e., not on word n but on word $n+1$) seems to agree with the idea of forced fixations in the E-Z Reader model and thus may be in line with serial word processing during reading, their temporal pattern (i.e., across all fixation durations on the target word) might be less intuitively explained by forced fixations. However, only simulations can show how the E-Z Reader model effectively behaves with respect to preview difficulty

[5] As SWIFT contains a very similar two-stage saccadic programming apparatus as compared with E-Z Reader, oculomotor explanations based on failed cancellation of saccade programmes during their non-labile stage, similar to the forced fixation account, could potentially also account for the relative difference between high- and low-frequency previews in the SWIFT model.

manipulations and how flexible the model becomes with respect to parafoveal processing effects because of its interaction of attention shifts and saccade programming.

At the same time, preview difficulty effects are just as well in line with parallel word-processing accounts during reading. Although current implementations of the SWIFT model (e.g., Schad & Engbert, 2012) do not adhere to such an interaction, the main idea of GDA models is that parafoveal and foveal processing somehow interacts with respect to the timing of the next eye movement. Future simulations with an accordingly modified SWIFT model would need to show whether parafoveal inhibition alone suffices to account for the spatio-temporal characteristics of the empirical preview difficulty effect. In contrast to the forced fixation account in E-Z Reader, which can be considered a by-product of the oculomotor system mainly independent of assumptions on how words are lexically processed, the implementation of parafoveal inhibition into SWIFT would test one of the core issues in the debate on serial versus parallel processing, which is the interaction (or cross-talk) between foveal and parafoveal (lexical) word processing during reading. s

2.6 Acknowledgements

We thank Liz Schotter, Denis Drieghe, Simon Liversedge, Hans Trukenbrod, and two anonymous reviewers for their valuable feedback on an earlier version of the manuscript. Data and R-Scripts are publicly available under DOI 10.17605/OSF.IO/KZ483 to replicate figures, tables, and statistical analyses reported in this article. The results have been presented at the 19th European Conference of Eye Movements (ECEM), Wuppertal, Germany.

2.7 Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

2.7.1 Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by Grants RI 2504/1-1 from the Deutsche Forschungsgemeinschaft.

2.8 Open Practices

The data from the present experiment are publicly available at the Open Science Framework website: DOI 10.17605/OSF.IO/KZ483

2.9 Supplemental material

Supplemental material for this article is available online.

Chapter 3

Bayesian parameter estimation for the SWIFT model of eye-movement control during reading

Stefan A. Seelig[a], Maximilian M. Rabe[a], Noa Malem-Shinitski[b],
Sarah Risse[a], Sebastian Reich[b,c,d], Ralf Engbert[a,c,d]

[a] Department of Psychology, University of Potsdam, Germany
[b] Institute of Mathematics, University of Potsdam, Germany
[c] Research Focus Cognitive Sciences, University of Potsdam, Germany
[d] Research Focus Data-Centric Sciences, University of Potsdam, Germany

Running Head: Bayesian parameter estimation

published 2020 in *Journal of Mathematical Psychology, 95*, 102313.
doi: 10.1016/j.jmp.2019.102313

Abstract

Process-oriented theories of cognition must be evaluated against time-ordered observations. Here we present a representative example for data assimilation of the SWIFT model, a dynamical model of the control of fixation positions and fixation durations during natural reading of single sentences. First, we develop and test an approximate likelihood function of the model, which is a combination of a spatial, pseudo-marginal likelihood and a temporal likelihood obtained by probability density approximation. Second, we implement a Bayesian approach to parameter inference using an adaptive Markov chain Monte Carlo procedure. Our results indicate that model parameters can be estimated reliably for individual subjects. We conclude that approximative Bayesian inference represents a considerable step forward for computational models of eye-movement control, where modeling of individual data on the basis of process-based dynamic models has not been possible so far.

3.1 Introduction

Dynamical models represent an important theoretical approach to cognitive systems, in particular, if we seek to explain time-ordered behavioral data such as sequences of movements. In dynamical models, sequential dependencies between observations are naturally explained by underlying dynamical principles that unfold over time when the model is simulated numerically (Beer, 2000; Van Gelder, 1998). Examples for the dynamical approach can be found in many fields of cognitive research, triggered by early examples from motor control (Erlhagen & Schöner, 2002; Haken, Kelso, & Bunz, 1985) or decision field theory (Busemeyer & Townsend, 1993).

Dynamical models generate highly specific predictions on sequential data that include statistical correlations between the subsequent observations over time. As a consequence, parameter inference for dynamical models must be carried out with the fully dynamical framework of *data assimilation* (Law, Stuart, & Zygalakis, 2015; Reich & Cotter, 2015). Here we investigate parameter inference in the SWIFT model of saccade generation during reading (Engbert et al., 2005), where the numerical computation of the model's *likelihood function* will be the fundamental concept and main contribution of this work.

In the research area of eye-movements during reading, a number of competitor models has been proposed. These models implement alternative assumptions on the interaction of word recognition and saccade generation (see Rayner & Reichle, 2010; Reichle et al., 2003, for overviews). However, there is currently a lack of quantitative model evaluations using objective concepts. First, due to the number of different effects in experimental data, models were often compared qualitatively: Does the model reproduce an experimentally-observed effect or not? Second, in complex cognitive models, parameters were mostly hand-selected or fitted based on minimization of an arbitrary loss-function that quantifies the difference between experimental and simulated data. Third, typical models could not be fitted to data from individual subjects so far. However, explaining interindividual differences is an important aspect of model evaluation, which is precluded when fitting procedures are data hungry and require pooling of data over a large number of participants. Since model identification and model comparison are general problems in psychological and cognitive sciences, Schütt et al. (2017) recently proposed a likelihood-based, statistically well-founded Bayesian framework for parameter estimation in cognitive models. We will demonstrate the feasability of this approach in the case of the SWIFT model for eye-movement control during reading.

In the following, the data assimilation framework will be applied to the SWIFT model of eye guidance in reading. The remaining part of this section consists of a short introduction to eye movement data and the specifics of likelihood functions for models of fixation sequences. In Section 2, we describe the details of the SWIFT model. A numerical approximation of the likelihood function is proposed and tested in Section 3. In Section 4, we use data from a set of readers to estimate SWIFT parameters and to model their interindividual differences. We close with a discussion of our results in Section 5.

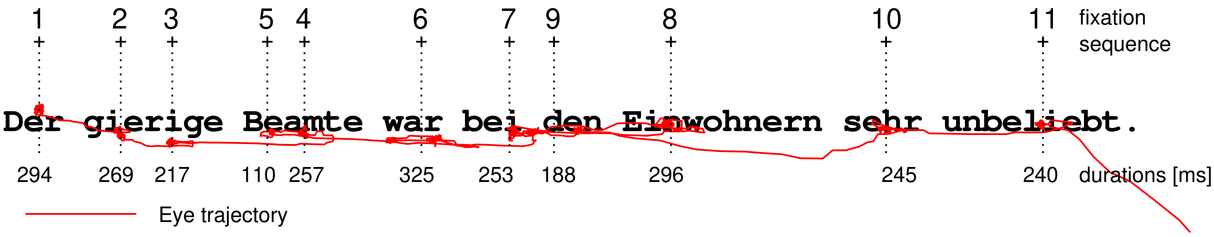

Figure 3.1: Sequence of fixations during reading. The scanpath indicates a series of fixations and saccades. Fixations are labeled by numbered dotted lines which indicate the horizontal positions. Numbers below the vertical lines are the corresponding fixation durations.

3.1.1 Eye-movement control during reading

Reading is based on successful word recognition, however, processing of words requires high-acuity vision that is confined to the center of the visual field (the fovea). Therefore, gaze shifts via fast eye movements (saccades) need to be generated to move words into the fovea for word identification. From this general behavioral pattern, reading may be looked upon as an important example of *active vision* (Findlay & Gilchrist, 2003), which is the notion that eye movements form an essential component for almost all visual perception.

When we read texts, we perform 3 to 4 saccades per second, resulting in fixations on different words with durations between 150 and 300 ms, on average. An example is presented in Figure 3.1, where 11 fixations are placed on the words of a given sentence. Fixation durations range from 110 ms to 325 ms. In this example, some words are fixated more than once. In the case of an immediate second saccade to the same word as the currently fixated word, the event is called a *refixation* (e.g., fixations 3, a forward refixation, and 5, a backward refixation). Some words are not fixated during first-pass reading, corresponding saccades are termed *skippings* (e.g., word 6, the article “den”, was skipped in *first-pass* reading). Furthermore, it happens in roughly 5 to 10% of the fixations that a corresponding saccade returns to a previously passed region of text, which are called *regressions* (e.g., when word 6, the previously skipped article, receives fixation 9). Taken together, only about 50% of the saccades are moving the gaze forward from word n to the next word $n+1$, which generates complicated *scanpaths* that are difficult to reproduce and predict by theoretical models of eye guidance during reading.

Eye movement research in reading has evolved into one of the fields of cognitive psychology that is strongly driven by computational models. Most of these models are based on simplified assumptions for several cognitive subsystems (e.g., oculomotor circuitry, attention and word recognition), while the core of the models is the orchestration of the subsystems to produce purposeful saccades for reading in a psychologically plausible framework. The way to this success has been paved by the E-Z Reader model (Reichle et al., 1998), a rule-based stochastic automaton model that is based on specific assumptions for the coupling of eye movements and visual attention. This model has been advanced over the years to include more and more specific assumptions (e.g., Reichle et al., 2009).

One of the major differences between existing models lies in the generation of different types of saccades (forward saccades, skippings, refixations and regressions). While some models make explicit assumptions on saccade types or are built to have internal states representing saccade types, an alternative model considered here is motivated by the dynamical field theory of movement preparation (S.-i. Amari, 1977; Erlhagen & Schöner, 2002), which communicates the aspiration to form a general framework for human motor control. The SWIFT[1] model (Engbert et al., 2002, 2005; Schad & Engbert, 2012) provides a coherent theoretical framework for reproducing all types of saccades that are observed during reading. Word processing maps to a distributed activation field that serves as a temporally evolving saccade targeting map. This model will be studied in detail with respect to parameter inference.

Given alternative theoretical models, model fitting and model comparisons will become an increasingly important topic in eye-movement research, as in cognitive science in general. So far, the minimization of ad-hoc statistical loss-functions has been used to obtain estimates for model parameters (e.g., Engbert et al., 2005; Reichle et al., 1998). For example, differences in word-frequency dependent distributions of fixation durations or skipping probabilities have been implemented as a measure of goodness-of-fit. We will replace these procedures by a Bayesian framework that exploits the likelihood function of the model.

Quantitative measures for eye movements during reading are characterized by strong interindividual differences (e.g., Risse, 2014). However, current computational models of eye-movement control could not reproduce and explain these obvious differences in human performance. It is a key message of the current work that the problem of modeling interindividual differences in reading using complex simulation models can be overcome when a likelihood-based framework of model identification, model parameter estimation, and model comparison is applied. We start with a discussion of the general concept of the likelihood function for dynamical cognitive models in the next section. The approximative computation of the likelihood function for the SWIFT model, which is the central contribution of the current work, is discussed in Section 3.

3.1.2 The likelihood function for dynamical cognitive models

The key theoretical concept for the current study is the likelihood function (see Myung, 2003, for a tutorial), which is a quantitative measure of the plausibility of an observation under the assumption of a specific model M. We assume that the model depends on a

[1]Saccade Generation With Inhibition by Foveal Targets

set of parameters $\boldsymbol{\theta}$ from parameter space $\boldsymbol{\Theta}$. In parameter inference, we are interested in the likelihood of the model parameter values $\boldsymbol{\theta}$ for model M given the experimental data,

$$L_M(\boldsymbol{\theta}|\text{data}) = P_M(\text{data}|\boldsymbol{\theta}) \ , \tag{3.1}$$

where $P_M(\text{data}|\boldsymbol{\theta})$ is the probability of the data given model M with parameters $\boldsymbol{\theta}$.

The maximum likelihood estimator $\hat{\boldsymbol{\theta}}_{ML}$ is obtained by maximization of the likelihood function, i.e.,

$$\hat{\boldsymbol{\theta}}_{ML} = \arg\max_{\boldsymbol{\theta}\in\Theta} L_M(\boldsymbol{\theta}|\text{data}) \ . \tag{3.2}$$

In mathematical models of eye-movement control, a model must be evaluated against a sequence of fixations. Thus, the data is a time-ordered sequence of fixations $F = \{f_i\}$, where each fixation f_i is characterized by a position x_i on the line of text, a fixation duration T_i, and, depending on the model, also a saccade duration s_i between fixation $i-1$ and fixation i.

In a dynamical model, fixation $f_i = (x_i, T_i, s_i)$ is generated from the sequence of previous fixations $f_1 \dots f_{i-1}$ under the control of the set of parameters $\boldsymbol{\theta}$ and, possibly, influenced by internal degrees of freedom $\boldsymbol{\xi}$, which will be discussed in Section 3.2.7. Since fixations are time-ordered, we can factorize the likelihood into a product of all fixations $i = 1, 2, ..., n$, which are found in the experimental fixation sequence $F = \{f_i\}_{i=1}^n$, i.e.,

$$\begin{aligned} L_M(\boldsymbol{\theta}|F) &= L_M(\boldsymbol{\theta}|f_1,\, f_2,\, \dots,\, f_n) \\ &= L_M(f_1|\boldsymbol{\theta}) \prod_{i=2}^{n} P_M(f_i|f_1,\, \dots,\, f_{i-1}, \boldsymbol{\theta}) \ , \end{aligned} \tag{3.3}$$

where $L_M(f_1|\boldsymbol{\theta})$ is the probability of the initial fixation starting at time $t = 0$. In typical experimental paradigms, however, this probability is one, since the experimental procedure determines the initial fixation position.

For complex cognitive models, the likelihood function can often be computed numerically. If numerical computation of the likelihood function is possible, we must be able to evaluate the likelihood for a large number of combinations of model parameter values $\boldsymbol{\theta}$ to find the maximum likelihood estimator, Eq. (3.2), based on a given fixation sequence F.

For the implementation of numerical computations, it is advantageous to compute the log-likelihood, given as

$$\begin{aligned} l_M(\boldsymbol{\theta}|F) &= \log(L_M(\boldsymbol{\theta}|F)) \\ &= \sum_{i=1}^{n} \log(L_M(f_i|f_1,\, \dots,\, f_{i-1}, \boldsymbol{\theta})) \ , \end{aligned} \tag{3.4}$$

which prevents the addition of very small numerical values that typically occur for some of the additive terms $L_M(f_i|f_1,\, \dots,\, f_{i-1}, \boldsymbol{\theta})$ for the fixations f_i.

If we can compute the log-likelihood $l_M(\boldsymbol{\theta}|F)$ for model M efficiently using numerical simulation, then it will be possible to apply Bayesian parameter inference (Gelman et al., 2013; Marin & Robert, 2007, for overviews). In Bayesian inference, we seek to compute the posterior distribution $P(\boldsymbol{\theta}|F)$ over the parameter vector $\boldsymbol{\theta}$ after the observation of the fixation sequence F. In addition to the likelihood that represents constraints from the experimental data, we specify a prior probability $Q(\boldsymbol{\theta})$ that indicates our a-priori knowledge on the model parameters. The posterior distribution is given by

$$P(\boldsymbol{\theta}|F) \propto Q(\boldsymbol{\theta}) L_M(\boldsymbol{\theta}|F) \;, \tag{3.5}$$

where the constant of proportionality, which is the normalization constant of the posterior, can be omitted, if Markov Chain Monte Carlo (MCMC) methods are used (Gilks et al., 1995; Robert & Casella, 2013).

So far, we discussed the structure of the likelihood function for a single experimentally observed fixation sequence F. In a typical experiment, however, we obtain a set of fixation sequences F_s from a participant who read a corpus of S sentences ($s = 1, 2, 3, ..., S$), i.e., the data set $\{F_s\}$ is composed of S fixation sequences. Since fixation sequences are statistically independent observations of the reading process, the numerical computation of the likelihood can be carried out independently for each fixation sequence F_s. This statistical independence can be exploited to accelerate computations via parallel evaluations of a large number of fixation sequences, which we will discuss in Section 3.3.

In summary, the likelihood function for dynamical models of sequential data factorizes as explained in Eq. (3.3), which turns out to be basis for incremental numerical computation. If we implement the computation in an efficient way numerically, then Bayesian parameter inference is available using MCMC methods. Before we discuss and apply the MCMC framework, we introduce the SWIFT model in the next section. In Section 3.2.7, we present the numerical computation of the likelihood function. The MCMC simulation for Bayesian inference will be discussed in Section 3.3.

3.2 The SWIFT model of saccade generation during reading

Since word recognition is the key process driving eye movements during reading, a natural assumption is that the time-course of ongoing word processing is closely linked to target selection for saccades. In the SWIFT model, each word is represented by a separate activation variable (lexical activation) that is tracking the word's current progress in word recognition. The resulting set of lexical activations determines the probability for saccade target selection (so-called spatial or *where* pathway). Whenever a saccade is prepared, the set of lexical activations provides a flexible mechanism for target selection. As time evolves, the relative activations change, so that a continuous-time representation of the next saccade target exists.

Fixation times are adjusted to the fixated (foveal) word by an inhibitory mechanism (the temporal or *when* pathway). According to an influential proposal (Findlay & Walker,

1999) the spatial and temporal pathways of saccade generation are partially independent. The SWIFT model is compatible with this view, in the sense that control of fixation duration and saccade target selection are basically independent, however, interactions exist due to the coupling of both pathways via the set of lexical activations.

3.2.1 Saccade target selection and temporal evolution of activations

Each word m in a sentence of N_w words is represented by a time-dependent activation $a_m(t)$. The activation is initially increasing during lexical access (word recognition), and later decreasing during post-lexical processing. The set of activations $\{a_j(t)\}$, $(j = 1, 2, 3, ..., N_w)$ must be built up by parallel processing of words, which is the key assumption that distinguishes SWIFT from other models (e.g., Engbert & Kliegl, 2011; Reichle et al., 2003). If a saccade target has to be selected at time t, then the probability $\pi_m(t)$ for target selection of word m is given by the relative activation, i.e.,

$$\pi(m,t) = \frac{(a_m(t))^\gamma}{\sum\limits_{j=1}^{N_w} (a_j(t))^\gamma} , \tag{3.6}$$

which is normalized as $\sum_{m=1}^{N_w} \pi_m(t) = 1$ for all $t > 0$. The parameter γ introduces a weighting of the set of lexical activations, so that switching between different selection schemes is controlled by a variation of γ:

$$\pi_m(t) \rightarrow \begin{cases} \text{winner-takes-all} & : \quad \gamma \rightarrow \infty \\ \text{Luce's choice rule} & : \quad \gamma = 1 \\ \text{random selection} & : \quad \gamma \rightarrow 0 \end{cases} . \tag{3.7}$$

An example for a simulated scanpath and the full time-series of lexical activation is illustrated in Figure 3.2. As one can see from figure, all internal sub-processes of the model are implemented by discrete random walks. In the leftmost column, the saccade timer increases as a one-step process from $n_1 = 0$ up to a maximum number N_t with transition rate w_1. The stepping rate was chosen as N_t/t_{sac}, so that the mean time to reach state N_t is the mean time inter-saccadic time t_{sac} of the model.

When the saccade timer terminates at state N_t, a new saccade timer run is initiated at $n_1 = 0$ and, at the same time, a labile saccade program start with $n_2 = 0$ until its threshold N_l is reached. If this labile program terminates, a saccade target is chosen (see asterisks in Fig. 3.2). After the non-labile stage, which is described by state variable n_3, the corresponding saccade (state variable n_4) is executed.

In addition to the saccadic processes, lexical activations are also described by discrete random walks (note, however, the increasing and decreasing parts in the case of lexical activations). Thus, all sub-processes saccade timing, labile and non-labile saccade programming, saccade execution, and change of lexical activations are represented as one-step stochastic processes between discrete states.

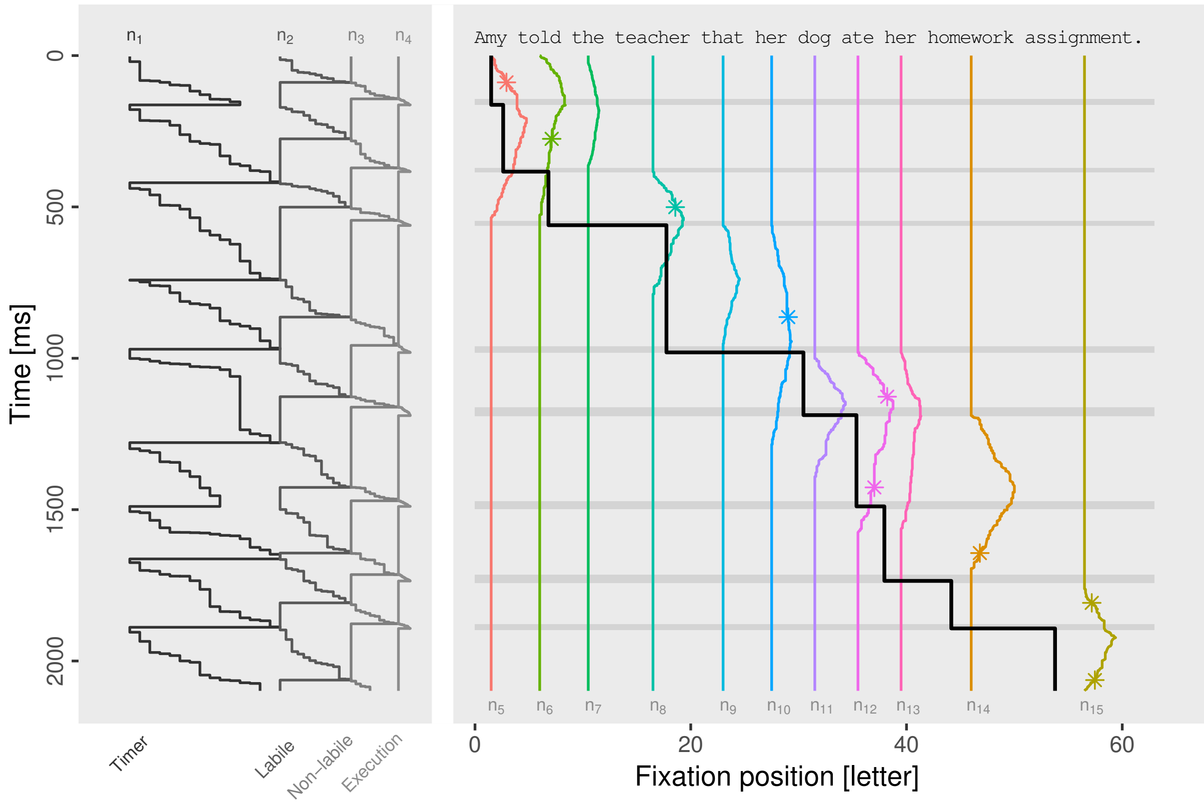

Figure 3.2: Simulation example for the SWIFT model. The activation field (colored lines) determines the target selection probability $\pi_m(t)$ that evolves dynamically over time (running downwards). The resulting scanpath (fixation sequence) is indicated by the black line. Several random walks (grey, left) generate saccade timer intervals and labile and non-labile saccade latencies. The transition between labile and non-labile stage is the point in time for saccade target selection (asterisk). The saccade timer sends commands to the saccade programming cascade, but also receives inhibition during foveal load (visible shortly after 1000 ms in the example) and is reset for refixations (e.g., second fixation).

The state of the model at time t is given by the vector $n = (n_1, n_2, ..., n_{4+N_w})$, where the components n_j represent the states of the subprocesses with transition rates w_j. Components 1 to 4 are saccade-related processes and additional stochastic variables n_5 to n_{4+N_w} are keeping track of the (post-)lexical processing of words. We assume a discrete-state, continuous-time stochastic process with Markov property, so that a one-step transition table describes all possible transitions between internal states (Tab. 3.1). In each of the possible transitions from state $n = (n_1, n_2, ...)$ to $n' = (n'_1, n'_2, ...)$ only one of the components n_i is changes by one unit, e.g., if the saccade timer generates a transition, then the model's internal change steps from $n = (n_1, n_2, n_3, ...)$ to $n' = (n_1 + 1, n_2, n_3, ...)$.

Table 3.1: Stochastic transitions between adjoined states from $n = (n_1, n_2, ...) \mapsto n' = (n'_1, n'_2, ...)$

Process	Transition to ...	Transition rate $W_{n'n}$
Saccade timer	$n'_1 = n_1 + 1$	$w_1 = N_t/t_{sac} \cdot (1 + h\, a_k(t)/\alpha)^{-1}$
Labile program	$n'_2 = n_2 + 1$	$w_2 = N_l/\tau_l$
Non-labile program	$n'_3 = n_3 + 1$	$w_3 = N_n/\tau_n$
Saccade execution	$n'_4 = n_4 + 1$	$w_4 = N_x/\tau_x$
Word processing	$n'_{4+j} = n_{4+j} \pm 1$	$w_{4+j} = N_a/\alpha \cdot \Lambda_j(t)$ (for word j)

A numerical algorithm for the simulation of a trajectory of the SWIFT model can be derived easily from our assumptions. The temporal evolution of the probability over the model's internal states is given by the master equation[2],

$$\frac{\partial}{\partial t} p(n,t|n'') = \sum_{n'} \left[W_{nn'} p(n',t|n'') - W_{n'n} p(n,t|n'') \right] \ , \tag{3.8}$$

which is specified by the transition probabilities $W_{n'n}$ for transitions between state vectors $n \mapsto n'$ shown in Table 3.1 with initial condition $p(n'', 0)$, the probability of state n'' at time $t = 0$. When simulating a single trajectory, the system is in a specific state n with certainty and the transition probabilities determine both the waiting time distribution for the next transition and the relative stepping probability to the adjoined states given in (Tab. 3.1), which will be explained below.

3.2.2 Temporal control of saccades and foveal inhibition

Gaze duration, defined as the sum of the durations of all immediately consecutive fixations on a word, is probably the best measure of required processing time for this word during natural reading (Rayner, 1998). Gaze durations and word recognition times depend linearly on the logarithm of the word's frequency (printed word frequency can be estimated from the word's occurrences in large text corpora). Since word recognition is the basis for text comprehension, an adaptive mechanism for the modulation of fixation duration by word frequency is essential for all models of eye-movement control.

In general, the required fixation duration for successful word recognition can be attained by two opposing mechanisms: The current fixation can be prolonged by inhibiting the next saccade or, alternatively, the word can be refixated to increase gaze duration. Experimentally, there is only a weak influence of word frequency on the mean first-fixation duration (Kliegl, Grabner, Rolfs, & Engbert, 2004). In contrast, we find a strong effect of

[2]The master equation can be interpreted as a conservation equation for probability (Gardiner, 1985; Van Kampen, 1992), where the temporal change of probability in state n on the left side of the equation equals the *gain* in probability for state n that is generated by transitions from neighboring states $n' \mapsto n$ and the *loss* in probability generated by transitions from n to neighboring states $n \mapsto n'$.

word frequency on the probability for refixation. Therefore, there is a preferred strategy for extending the processing time (gaze duration) via generation of a refixation. However, saccade-inhibiting processes can be assumed to contribute a weaker effect (compared to refixation) to the increase in gaze duration by prolonging the ongoing fixation (Engbert et al., 2002, 2005).

Motivated by these observations, the second central assumption in the SWIFT model is *random timing* of fixation duration with additional *foveal inhibition* (Engbert et al., 2002) that delays the start of the next saccade program to extend the current fixation. We assume that foveal inhibition modulates the transition rate w_1 for transitions between elementary steps of a random-walk that implements the saccade timer (leftmost column in Fig. 3.2), i.e.,

$$w_1 = \frac{N_t}{t_{\text{sac}}} \cdot \left(1 + \frac{h}{\alpha} a_k(t)\right)^{-1} , \tag{3.9}$$

where N_t is the number of states of the timer's random walk and t_{sac} is the mean value of the timer; the activation $a_k(t)$ of the fixated word k (i.e., the word in the fovea) at time t is the key variable that modulates the transition rate of the timer. Using numerical simulations of the model, it can be shown that for $h > 0$, foveal inhibition can produce a modulation of the fixation duration that is in good agreement with experimental data (Engbert et al., 2002, 2005).

3.2.3 Character-based visual processing

Word recognition starts with visual processing of letters, which is done in parallel for all the letters of a given word. We define the spatial region where word activations can be influenced in the model as the *processing span*. Within this region, parallel processing is limited by the fact that processing rate depends on the letter's *eccentricity* (i.e., the distance of the letter position from the position of the current fixation). Mathematically, we define an inverted parabolic processing span from the fovea to position $-\delta_L$ on the left and to position $+\delta_R$ on the right of fixation, i.e.,

$$\lambda(\epsilon) = \lambda_0 \cdot \begin{cases} 0\,, & \text{for} \quad \epsilon < -\delta_L \\ 1 - \epsilon^2/\delta_L^2\,, & \text{for} \quad -\delta_L \le \epsilon < 0 \\ 1 - \epsilon^2/\delta_R^2\,, & \text{for} \quad 0 \le \epsilon \le \delta_R \\ 0\,, & \text{for} \quad \delta_R < \epsilon \end{cases} , \tag{3.10}$$

where λ_0 is a constant given as

$$\lambda_0 = \frac{3}{2} \cdot \frac{1}{\delta_L + \delta_R} , \tag{3.11}$$

which is necessary to normalize the total processing rate, i.e., $\int_{-\infty}^{+\infty} \lambda(\epsilon) \mathrm{d}\epsilon = 1$.

Experimentally, a strong asymmetry of the *perceptual span* with an extension of 4 to 5 letters to the left of the fixation position and up to 15 letters to the right was found (Rayner et al., 1980). Therefore, parameters δ_L and δ_R should be estimated separately from experimental data. In the following, we estimate $\delta_0 \equiv \delta_L = \delta_R$ for simplicity.

3.2.4 Word-based processing rate

Because of the assumption of a processing span, Eq. (3.10), processing rates for letters depend on spatial eccentricities. Letter j of word i is processed with rate $\lambda(\epsilon_{ij})$, if it is located at a spatial position with eccentricity $\epsilon_{ij}(t)$ relative to gaze position at time t. This letter-based processing rate must be related to the effective word-based processing rate $\Lambda_i(t)$ of word i at time t.

Because of parallel processing of the letters of a given word, each letter contributes to word recognition. In the case of long words, some letters will have large eccentricities, so that their processing rate will be small (or zero) according to Eq. (3.10). To capture these opposing effects in a parametric model, we make the assumption that the word-based processing rate has the form

$$\Lambda_i(t) = M_i^{-\eta} \sum_{j=1}^{L_i} \lambda(\epsilon_{ij}(t)) \; , \tag{3.12}$$

where M_i is the word length (i.e., number of letters) of word i and η is the word length exponent, with $0 < \eta < 1$. For $\eta = 0$, long words will have a processing advantage. For $\eta = 1$, word processing rate is the arithmetic mean of the letter-based processing rates (mean over all letters of a given word); therefore, we will observe a disadvantage for long words in the case $\eta = 1$. We expect a numerical value for η about 0.5.

With the assumptions on spatial aspects of letter- and word-based processing rates, the temporal aspects of word processing need to be specified. As discussed for the motivation of the SWIFT model, a time-dependent activation field will provide probabilistic control of saccadic eye movements. Word-based activations $a_i(t)$ for the words of a given sentence are increasing during the initial stage of processing called *lexical processing*. After reaching the maximum of activation D_i for word i, the activation starts to decrease (*post-lexical processing*). The maximum of activation is interpreted as processing difficulty, which is a logarithmic function of word frequency Ω_i, i.e.,

$$D_i = \alpha \left(1 - \beta \frac{\log \Omega_i}{\log \Omega^{\max}} \right) \; , \tag{3.13}$$

where $\Omega^{\max}$ is the highest word frequency in a given language and parameter β determines the strength of the word frequency effect.

For word processing, we assume that current activation for each word $i = 1, 2, 3, ..., N_w$ is related to the discrete state n_{4+i} of word processing (Tab. 3.1), given by

$$a_i(t) = D_i \frac{n_{4+i}}{N_a} \; , \tag{3.14}$$

where $D_i$is the word's processing difficulty, Eq. (3.13).

Global decay of activation. Maintaining words in working memory during reading cannot be done without loss. Since word activations $\{a_n(t)\}$ represent the state-of-processing, we introduce a global decay of activation. If the processing rate of a word is smaller than the constant ω, then we assume a decay of activation with rate ω.

Processing during saccades. During saccadic eye movements, lexical processing is paused because of *saccadic suppression* (Matin, 1974). In the SWIFT model, lexical processing is paused during saccades in the lexical processing stage (increasing activation), while post-lexical processing (decreasing activation) continues during saccades.

3.2.5 Oculomotor assumptions

Our assumption of two-stage saccade programming are motivated by the experimental findings of the double-step paradigm (Becker & Jürgens, 1979). A saccade program starts with a labile stage; during this stage, the saccadic gaze center is forced to prepare the next saccade (Findlay & Walker, 1999), however, a new decision to start a labile saccade program during an ongoing labile stage leads to cancelation and replacement of the earlier saccade program. After the transition to the non-labile stage, the saccade can no longer be canceled or modified.

Oculomotor errors make an important contribution to eye-movement control during reading. In 1988, based on the analysis of initial fixation positions within words, McConkie and coworkers suggested that a considerable fraction of saccades landed on different words than the intended target words (McConkie et al., 1988). Using an iterative oculomotor modeling approach, Engbert and Nuthmann (2008) showed that about 10% to 20% of the saccades during natural text reading are mislocated on an unintended word.

McConkie et al. (1988) showed that saccadic errors can be decomposed into a random (approximately Gaussian) error component and a systematic shift (called saccadic range error). The critical variable that determines the size of both random and systematic error components turned out to be the intended saccade length (distance d from the launch site of the saccade to the center of the target word). Therefore, saccades targeting a word center at $x = 0$ will be normally distributed with

$$x \sim \mathcal{N}(\epsilon_{\text{sre}}, \sigma^2_{\text{sre}}) \ , \tag{3.15}$$

where both parameters depend linearly on the intended saccade length d, i.e.,

$$\epsilon_{\text{sre}} = r_1 - r_2\, d \tag{3.16}$$

$$\sigma_{\text{sre}} = s_1 + s_2\, d \ , \tag{3.17}$$

where d is the physical distance between the launch site of the saccade and the word center of the target word, measured in units of character spaces. The oculomotor parameters r_1, r_2, s_1, and s_2 will vary depending on the type of saccade (e.g., refixation or skipping), which is discussed in earlier papers (Engbert et al., 2005; Krügel & Engbert, 2010). We would like to remark that McConkie et al.'s descriptive model of saccadic errors could be replaced by a process-oriented Bayesian model (Engbert & Krügel, 2010; Krügel & Engbert, 2014) in perspective.

Modulation of the duration of the labile stage. An important problem is the observation of a reduced average fixation duration for refixations. As a solution, we assume

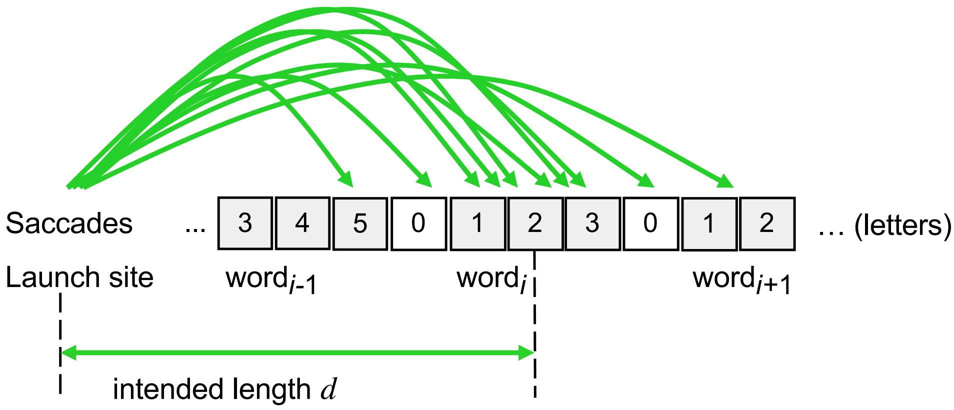

Figure 3.3: Saccades start at a launch site and aim at the word center of the selected target word i. Oculomotor errors are normally distributed, which can lead to misplaced fixations on word $i-1$ (undershoot error) or word $i+1$ (overshoot error). Both the standard deviation σ_{sre} and the mean shift ϵ_{sre} from the intended word's center depend on the intended saccade length d.

that the duration of the labile stage of saccade programming is reduced by factor R ($0 < R \leq 1$), if the fixation is a refixation.

Closely related is the phenomenon of mislocated fixations (Engbert & Nuthmann, 2008). If the realized fixation position (the saccadic landing position) strongly deviates from the word center, so that the landing position will fall onto the neighboring word, then a mislocated fixation will occur. In this case, the duration of the next saccade program will be reduced by factor M ($0 < M \leq 1$). Such a mechanism is a possible explanation of the inverted optimal viewing-position effect (Nuthmann et al., 2005; Vitu et al., 2001) of fixation durations that indicates reduced average fixation duration at word edges compared to the word center. In the SWIFT version used here, the probability of misplaced fixation is given as $p_{mis} = 0.9 \cdot (2\delta/M)^4$, where δ is the fixation error (distance from word center) and M is the length of the fixated word.

3.2.6 Numerical simulation and model parameters

For numerical simulations of single trajectories of the SWIFT model, the *minimal process method* by Gillespie (1976), an exact and efficient numerical algorithm, can be derived from the master equation, Eq. (3.8). If the model is in state n at time $t_0 = 0$ with certainty, all other states will have zero probability, i.e., $p(n', t|n)$ for $n' \neq n$. Therefore, the master equation, Eq. (3.8), reduces to

$$\frac{\partial}{\partial t} p(n,t|n) = -\sum_{n'} W_{n'n}\, p(n,t|n) = -W_n\ p(n,t|n)\ , \tag{3.18}$$

where $W_n = \sum_{n'} W_{n'n}$ is the total transition probability from state n. From Equation (3.18), we obtain an exponentially distributed waiting time for the next transition from

Table 3.2: Model parameters of the SWIFT model. Numerical values are chosen in agreement with earlier publications (see text).

Parameter	Symbol	Typical Value	Reference
Lexical difficulty: Intercept	α	50	Eq. (3.13)
Lexical difficulty: Slope	β	0.75	Eq. (3.13)
Processing span	$\delta_0 = \delta_{L,R}$	8	Eq. (3.10)
Word-length exponent	η	0.5	Eq. (3.12)
Saccade timer	t_{sac}	250 ms	Tab. (3.1)
Foveal inhibition	h	0.6	Eq. (3.9)
Labile saccade program	τ_l	120 ms	Tab. (3.1)
Non-labile program	τ_n	80 ms	Tab. (3.1)
Saccade execuation	τ_x	20 ms	Tab. (3.1)
Refixation factor	R	0.9	Sec. 3.2.5
Mislocated fixation	M	1.5	Sec. 3.2.5

state n to an adjoined state $n' \neq n$. Following Gillespie (1976), a two-step algorithm can be derived: In step 1, an exponentially-distributed random number is generated; in step 2, a transition (Tab. 3.1) is chosen according to relative transition probabilities, $W_{n'n}/W_n$ with $n' \neq n$. This algorithm is numerically efficient, since it restricts computations to the transitions when simulating the system's trajectory.

For the simulations in this paper we used a restricted version of the SWIFT model to reduce the number of free parameters to 11 (Tab. 3.2; see Engbert et al., 2005). Moreover, we fixed seven of these parameters to estimate four free parameters in the simulation examples. Future simulation studies will be carried out with more free parameters (see Sec. 3.4). The number of possible random-walk states varies between subprocesses; based on earlier simulations (Schad & Engbert, 2012), we used the following numbers: $N_t = 15$ (saccade timer), $N_l = 12$ (labile saccade stage), $N_n = 10$ (non-labile saccade stage), $N_x = 20$ (saccade execution), and $N_a = 30$ (word activations).

3.2.7 Likelihood function for the SWIFT model

For the parameter estimation procedure discussed in the introduction, we aim at a framework that computes the likelihood of a series of experimentally observed fixations incrementally, Eq. (3.3). For fixation f_i, we need to compute the likelihood function $L_M(f_i|f_1, \ldots, f_{i-1}, \boldsymbol{\theta}, \boldsymbol{\xi})$ given the previous fixations $f_1, f_2, ..., f_{i-1}$, the model parameters $\boldsymbol{\theta}$, and the internal states $\boldsymbol{\xi}$ of model M, which were not addressed in Eq. (3.3). In SWIFT each fixation event $f_i = (x_i, T_i, s_i)$ is defined by a fixation position x_i given by the fixated word v_i and the fixated letter l_i within the word, the fixation duration T_i, and the saccade duration s_i. The likelihood for fixation f_i is composed of a spatial contribution and a temporal contribution. At time t, fixation i starts on letter l_i of word v_i, which is predicted by the SWIFT model with a probability determined by word activations and oculomotor assumptions. After fixation i started, the model can make another prediction for the fixation duration T_i of fixation i. Next, the likelihood for fixation i can be decomposed into the spatial and temporal contributions, i.e.,

$$L_M(v_i, l_i, T_i | F_{i-1}, \boldsymbol{\theta}, \boldsymbol{\xi}) = L_{\text{temp}}(T_i | v_i, l_i, F_{i-1}, \boldsymbol{\theta}, \boldsymbol{\xi}) \cdot L_{\text{spat}}(v_i, l_i | F_{i-1}, \boldsymbol{\theta}, \boldsymbol{\xi}) \;, \tag{3.19}$$

where we introduced $F_{i-1} \equiv \{f_1, f_2, \ldots, f_{i-1}\}$ to simplify the notation.

For the *spatial likelihood* L_{spat}, the dynamically evolving word activations in SWIFT determine the time-dependent probability for selecting a particular word as the next target word. Additionally, the target-selection probability is modified by oculomotor noise. Due to dynamical dependencies, we compute the likelihood of an experimentally realized fixation position based on the previous fixations. However, the internal states $\boldsymbol{\xi}$ are given by the stochastic dynamics and are, therefore, unknown. In principle, we could integrate over many possible realizations of the internal states $\boldsymbol{\xi}$, which is, however, time-consuming for the numerical computations. Therefore, we compute L_{spat} for one realization of the internal states $\boldsymbol{\xi}$, which results in fluctuating numerical values for L_{spat}. Thus, instead of integrating out the internal degrees of freedom $\boldsymbol{\xi}$, we used a pseudo-marginal likelihood (Andrieu & Roberts, 2009) and eliminated the dependence on $\boldsymbol{\xi}$ for the spatial likelihood in Eq. (3.19).

For the *temporal likelihood* L_{temp}, SWIFT computations start with a realized fixation position on letter l_i of word v_i, however, with internal states $\boldsymbol{\xi}$. Given this fixation position, the distribution of fixation durations can be predicted by the model. The generated estimate of the likelihood of the experimentally realized fixation duration is approximated by averaging over many realizations of the internal states $\boldsymbol{\xi}$ (e.g., the internal states of the various saccade programming stages). As a result, both L_{temp} and L_{spat} are random variables, which will be discussed in detail in the next two sections.

3.2.8 Spatial likelihood

In SWIFT, saccadic gaze shifts are generated in two steps: First, a target word is determined in a probabilistic selection process based on relative word activations. Second, after a short delay generated by saccade programming, the saccade is executed with oculomotor errors influenced by the saccadic landing position distribution. These oculomotor errors induce stochastic variability in the within-word fixation position and can also induce mislocated fixations (Engbert & Nuthmann, 2008; Nuthmann et al., 2005), where the realized fixation position is placed on a different word than the selected target.

The combination of activation-based saccadic selection and oculomotor errors generates a non-zero probability for all fixation positions (Fig. 3.3). The target selection probability $\pi(m, t - \tau_n - \tau_x)$ (see. Eq. 3.6) is the probability of selecting word m as a saccade target for a fixation starting at time t. It is important to note that target selection occurs at the time of transition from the labile to the non-labile saccade program, so that the probability $\pi(.)$ for selecting the next target word has to be evaluated with an average time delay $\tau_n + \tau_x$. According to our oculomotor assumptions, the saccadic error generates a probability $q(v, l|m, x_{\text{gaze}})$ of fixating word v at letter l given that word m is the selected target word and x_{gaze} is the previous gaze position (or saccade launch site). Thus, the spatial likelihood of an observed saccade starting at time t_i towards letter position l of word v is therefore given by

$$L_{\text{spat}}(v, l|F_{i-1}, \boldsymbol{\theta}) = \sum_{m=1}^{N_w} \pi(m, t_i - \tau_n - \tau_x)\, q(v, l|m, x_{\text{gaze}}) \;, \tag{3.20}$$

where we dropped the conditional arguments to simplify the notation. Moreover, the time-dependency is now written explicitly, since t_i for the computation of the spatial likelihood of fixation i is given by the sum of fixation durations and saccade durations of the previous fixations in the sequence, $t_i = \sum_{l=1}^{i-1} T_l + s_l$.

The oculomotor system generates systematic and random errors that introduce deviations between the target word's center and the realized fixation position. In SWIFT, we adopt McConkie et al.'s (1988) range-error framework by assuming a Gaussian distribution that is shifted with respect to the target word's center. Thus, the probability of landing at letter l of word v, given a target word m, is given by

$$q(v, l|m, x_{\text{gaze}}) = \frac{1}{\sqrt{2\pi}\sigma_{\text{sre}}} \exp\left(-\frac{((v_m + \epsilon_{\text{sre}}) - x_{n,l})^2}{2\sigma_{\text{sre}}^2}\right) \cdot \Delta x \;, \tag{3.21}$$

where v_m is the spatial position of the target word's center, $x_{v,l}$ is the spatial position of the fixated letter l of word v, and $\Delta x = 1$ is the unit width of a letter. The oculomotor parameters $\epsilon_{\text{sre}}(d)$ and $\sigma_{\text{sre}}(d)$ of the range-error model specify systematic shift (saccadic range error) and standard deviation of the random error (oculomotor noise), respectively, Eqs. (3.16, 3.17); the intended saccade length $d = \|v_m - x_{\text{gaze}}\|$ is given as the distance between the target word's center v_m and the fixation position before the saccade x_{gaze}.

3.2.9 Temporal likelihood

Because of two-stage saccade programming and due to the fact that fixations are bounded by two saccades in time, SWIFT's fixation durations are given as linear combinations of realizations of random variables. For the saccade timer and saccade programming stages, resulting durations are gamma-distributed random variables, which are generated by continuous-time discrete-state random walks according to the master equation, Eq. (3.8).

The saccade timer controls the initiation of the saccade programming cascade with consecutive labile and nonlabile stages and a saccade execution stage. The time interval between the end point of the previous and the beginning of the next saccade execution is defined as the experimentally observed fixation duration. However, the saccade timer is continuously inhibited by word activations. As a consequence, the mean waiting times (the inverse of the transition probabilities) of the elementary steps of the saccade timer's random walk will be time-dependent. Additionally, the mean durations of the labile stages of saccade programming depend on the type of fixation (i.e., whether it is a refixation, a mislocated fixation, or neither of these). Finally, if the saccade timer produces a short interval, then saccade cancelation will be likely, which results in a higher mean value of the predicted fixation duration.

Since each fixation duration is bounded by two saccades (i.e., the ith fixation duration lies between $(i-1)$th saccade offset and ith saccade onset), each observed fixation duration T_i is compared to the simulated realization $\tilde{T}_i$ that is given as the sum of the following terms (see Fig. 3.4a),

$$\tilde{T}_i = \tilde{c}_i + \tilde{\tau}_i^l + \tilde{\tau}_i^n - \tilde{\tau}_{i-1}^l - \tilde{\tau}_{i-1}^n - \tilde{\tau}_{i-1}^x \ , \tag{3.22}$$

where $\tilde{c}_i$ is the realized saccade timer duration, $\tilde{\tau}_i^l$ and $\tilde{\tau}_i^n$ are realized durations of the labile and non-labile saccade programming stages respectively, and $\tilde{\tau}_i^x$ is the realized saccade duration.

Our strategy for the computation of the temporal likelihood of the ith fixation duration T_i is to simulate many realizations of $\tilde{T}_i$ from Eq. (3.22) to numerically approximate the theoretical distribution of fixation durations with kernel density estimation[3]. In the context of Bayesian analysis, this approach is termed *probability density approximation* (PDA) method (Holmes, 2015; Palestro et al., 2018; Turner & Sederberg, 2014), which falls into the broad class of likelihood-free procedures in *approximate Bayesian computation* (ABC; see Sisson & Fan, 2011, for a review).

Since all of the terms in Eq. (3.22) are random realizations of stochastic variables, the order of terminations of the subprocesses shown in Fig. 3.4(a) can be violated. In the following, we discuss all possible cases:

1. *Labile pausing* happens if the labile saccade program terminates during an ongoing non-labile saccade program. Since we assume that there cannot be more than one non-labile saccade program active at a time, the current labile program is paused immediately before termination, thus its duration is extended until the current non-labile program and saccade execution finish (Fig. 3.4b). Formally, this situation is encountered if $\tilde{c}_i + \tilde{\tau}_i^l < \tilde{\tau}_{i-1}^l + \tilde{\tau}_{i-1}^n + \tilde{\tau}_{i-1}^x$. In this case, the interval $\tilde{\tau}_i^l$ is increased and the calculation of $\tilde{T}_i$ is simplified to the duration of the non-labile saccade program, i.e.,
$$\tilde{T}_i = \tilde{\tau}_i^n \ . \tag{3.23}$$
Since the duration of the labile program is extended, however, there will be increased probability for the saccade timer to terminate during the ongoing labile program, while will cause saccade cancelation.

[3]While it is possible to derive an iterative algorithm for the distribution of linear combinations of gamma-distributed random numbers (S. V. Amari & Misra, 1997; Coelho, 1998), it turned out that these solutions are numerically unstable.

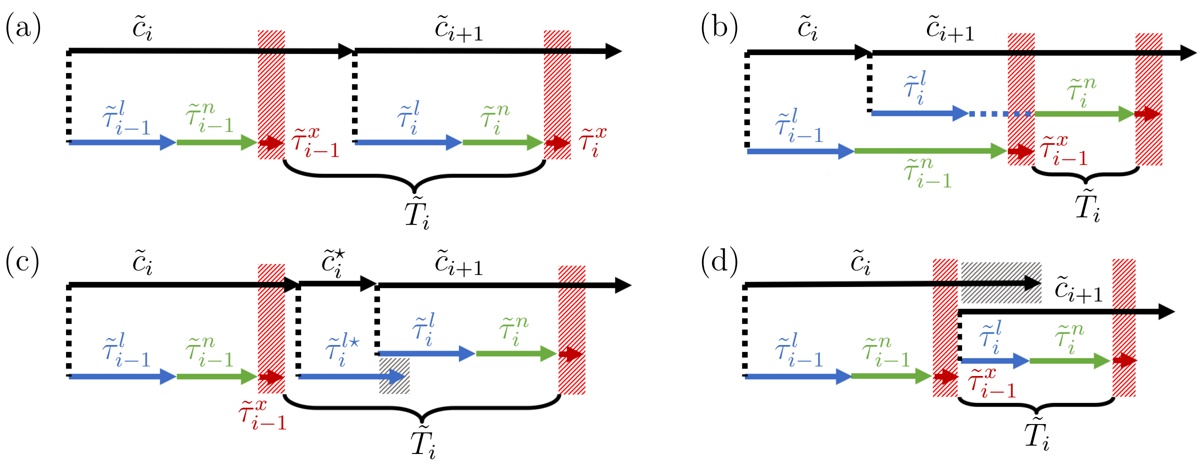

Figure 3.4: Schematic illustrations of the generation of fixation durations for different types of fixations in SWIFT. (a) *Standard case*: The fixation duration is calculated from the difference of the sum of the saccade timer c_i, the labile and nonlabile saccade latencies $\tilde{\tau}^l_i$ and $\tilde{\tau}^n_i$, respectively, and the sum of saccade latencies $\tilde{\tau}^l_{i-1}$, $\tilde{\tau}^n_{i-1}$ and $\tilde{\tau}^l_{i-1}$. (b) *Labile pausing*: If a saccade program reached the non-labile stage it cannot be aborted anymore. A newly started labile programming stage will transition to its non-labile stage only after the current saccade program is terminated at saccade offset. (c) *Saccade cancelation*: If the saccade timer finishes earlier than the concurrent labile saccade program, the ongoing labile saccade program is canceled—consequently, both the labile program and the saccade timer are restarted. The realized duration of the premature saccade timer $\tilde{c}^*_i$ is added to the new realization $\tilde{c}_i$. (d) *Refixation and Mislocated Fixation*: If the current fixation is either a refixation or considered to be a mislocated fixation, the saccade timer realization $\tilde{c}_i$ is reset immediately at fixation onset and a new labile saccade program is initiated. The fixation duration is then given as the sum of the current labile and non-labile durations $\tilde{\tau}^l_i$ and $\tilde{\tau}^n_i$ respectively.

2. *Saccade cancelation* occurs if the main saccade timer realization $\tilde{c}_{i+1}$ terminates during an ongoing labile saccade programming stage $\tilde{\tau}^l_i$, i.e., $\tilde{c}^\star_i < \tilde{\tau}^{l\star}_i$, which is illustrated in Figure 3.4c. In this case the labile saccade program is canceled and replaced with the new labile saccade program initiated by restarting of the saccade timer. As a result, the duration of the timer $\tilde{c}_i$ in Eq. (3.22) is replaced by the sum $\tilde{c}_i + \tilde{c}^\star_i$. Therefore, the corresponding distribution T_i for saccade cancelation is given by

$$\tilde{T}_i = \tilde{c}_i + \tilde{c}^\star_i + \tilde{\tau}^l_i + \tilde{\tau}^n_i - \tilde{\tau}^l_{i-1} - \tilde{\tau}^n_{i-1} - \tilde{\tau}^x_{i-1}\,, \qquad \text{if} \quad c^\star_i < \tilde{\tau}^{l\star}_i\,. \tag{3.24}$$

In principle, saccade cancelation can happen repeatedly within the same fixation, depending on the choice of parameters.

3. *Refixations and mislocated fixations* represent another special case, where a new saccade program is triggered immediately after the fixation onset (Fig. 3.4d). In

both cases the saccade timer realization $\tilde{c}_i$ is reset and a new labile saccade program is initiated. The mean duration of the new labile stage is modified by coefficients $f^r = 1/R$ and $f^m = 1/M$ for refixations and mislocated fixation, resp. (see 3.2.5). As a result, the observed fixation duration is given as

$$\tilde{T}_i = f^{r,m}\tilde{\tau}_i^l + \tilde{\tau}_i^n \ . \tag{3.25}$$

The SWIFT model includes inhibition of fixation durations by word activation; in its simplest form, the activation of the fixated (foveal) word inhibits the fixation duration by decreasing the transition rates of the saccade timer (Eq. 3.9). Because of the complicated time-course of the activation field (i.e., sudden changes of activation evolution due to saccades), stochastic simulations are necessary to estimate the distribution of $\tilde{T}_i$.

To compute the likelihood $L_{\text{temp}}(T_i)$ of an observed fixation duration T_i we first simulate the activation evolution for words in the perceptual span from time $t = 0$ until the point in time that corresponds to the end of fixation i. We start simulating the stochastic contributions by initially going backwards from the time of fixation onset by sampling the saccade latencies $\tilde{\tau}_{i-1}^x$, $\tilde{\tau}_{i-1}^n$, and $\tilde{\tau}_{i-1}^l$ to determine the onset of the saccade timer c_i. The previously sampled activations provide information for the simulation of the saccade timer with inhibition by foveal word activations, similar to the generative process. If $\tilde{c}_i < \tilde{\tau}_{i-1}^l$, both realizations are discarded and sampled again with the same procedure (we are not interested in saccade cancelation events which do not affect the fixation duration under consideration). The offset of $\tilde{c}_i$ demarks the onset of $\tilde{c}_{i+1}$ and, following the rules of the previously discussed order violations, we can easily simulate the timer cascade until fixation offset and hence obtain a sample from the distribution of fixation durations as provided by the SWIFT framework with respect to the history of the fixation sequence.

Once $N = 300$ fixation durations are sampled, the distribution of T_i^{exp} is approximated via KDE. Increasing the number of samples increases the accuracy of the approximation but is costly in terms of computation time. For the density estimation we use the Epanechnikov kernel (Epanechnikov, 1969) with a bandwidth setting according to Scott's rule (Scott, 2015). The Epanechnikov kernel is computationally efficient as it only integrates samples within its limited interval given by the bandwidth. However this can result in situations where no data point is covered by the kernel. To prevent estimates with zero probability, the bandwidth of the kernel was adjusted to the 1.1-fold of the distance between T_i^{exp} and the nearest sample of $\tilde{T}_i$, so that at least one sample will lie within the kernel.

3.2.10 Evaluation of the log-likelihood using single-parameter variations

A simple test of the likelihood function and its inherent stochastic contributions can be done by repeatedly evaluating the likelihood of a simulated dataset for which the parameters are known and keeping all parameters but one at their respective true values

Table 3.3: Parameters of the SWIFT model considered in Bayesian estimation; true values apply to the synthetic data generated for verification of the likelihood function.

Parameter	Symbol	Range	True value
Saccadic timer	t_{sac}	150 ... 350 ms	260 ms
Refixation factor	R	0.2 ... 1.8	0.9
Processing span	δ_0	4 ... 15	8.5
Word length exponent	η	0 ... 1	0.4

(i.e., the values used in generating the data). Systematically varying the parameter under consideration reveals its impact on the likelihood. Since the likelihood function is composed of two terms from spatial and temporal contributions (Eq. 3.19), separating both components can also prove insightful with regard to the strength and direction of the parameter's influence.

To investigate the properties of the likelihood function for a relevant subset of parameters, we simulated 1624 fixations on 114 sentences (Fig. 3.5) from the sentence corpus of Risse and Seelig (2019). The examined parameters are given in Table 3.3, with the remaining parameters set according to Table 3.2. The likelihood was then evaluated for 1000 different, evenly spaced values within the given interval (Table 3.3) separately for each parameter. Since all other parameters were fixed at their true values, any systematic change in the resulting log-likelihood can only be attributed to the parameter under consideration.

Figure 3.5a indicates that the saccade timer t_{sac} influences the temporal likelihood, while there is no influence on the spatial likelihood. A similar behavior is observed for the refixation factor R (Fig. 3.5b). In both cases, there is a clear maximum in the likelihood profile at the true parameter values, $t_{\mathrm{sac}} = 260$ ms and $R = 0.9$, resp. A different dependence can be seen for the processing span δ_0, which clearly influences the spatial likelihood (maximum at the true value $\delta_0 = 8.5$), but exerts only a minimal influence on the temporal likelihood (Fig. 3.5c). For the word-length exponent η, there is an influence on both spatial and temporal likelihoods (Fig. 3.5d), with a maximum for both likelihood profiles at the true parameter value $\eta = 0.4$.

Thus, our numerical implementation of the likelihood function indicates clear maxima at the true parameter values for simulated data, while stochastic fluctuations due to the approximative account for internal degrees of freedom $\boldsymbol{\xi}$ are small. In the next section, we will apply an adaptive MCMC framework for Bayesian parameter estimation using simulated and real (experimental) data.

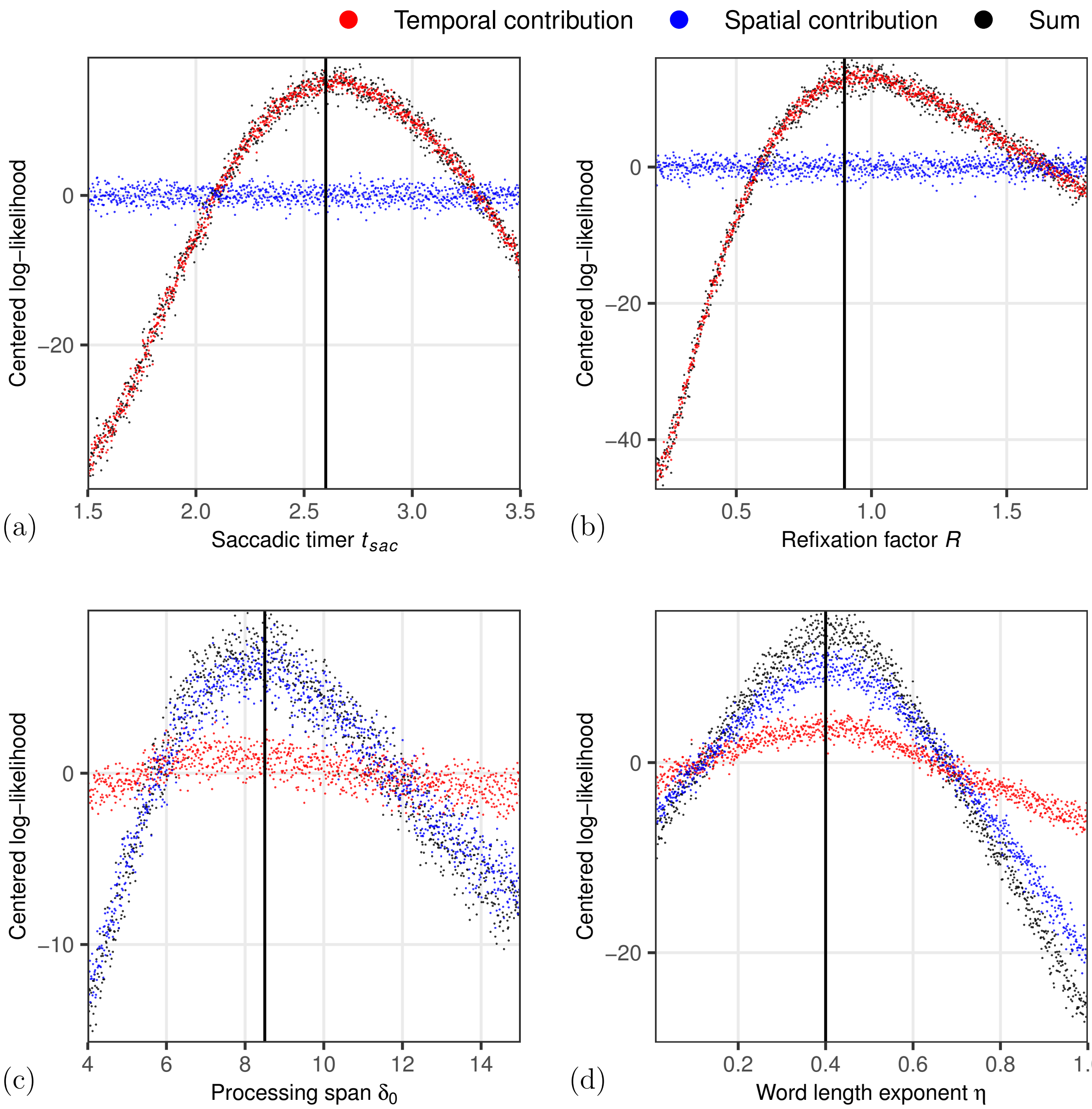

Figure 3.5: Temporal (red) and spatial (blue) contributions to the total (black) log-likelihoods of a simulated dataset (1624 fixations on 114 sentences from the corpus of Risse and Seelig (2019)). Single parameters were varied within an interval around the respective true parameter value used in creating the data. The log-likelihoods were centered around their respective mean value.

3.3 Likelihood-based parameter inference using MCMC

With the implementation of the numerical computation of the likelihood function for the SWIFT model from the previous section, we developed the critical step for adopting the Bayesian framework for parameter inference. We will discuss the Markov Chain Monte Carlo approach used for inference, discuss the efficient implementation on a digital computer, present results for parameter recovery from simulated data with known parameters, and, finally, estimate parameters for experimental data.

3.3.1 Markov Chain Monte Carlo simulation for the SWIFT model

As described in Section 3.1.2, the computability of the likelihood $L_M(\boldsymbol{\theta}|F)$, Eq. (3.3), for a given set of parameters $\boldsymbol{\theta}$ and a given fixation sequence F is critical for maximum-likelihood and Bayesian inference. For the numerical procedures of Markov Chain Monte Carlo type, we use a variant of the Metropolis Hastings (MH) algorithm (Hastings, 1970). In the random-walk MH algorithm, a random walk in the parameter space is generated, where the probability of the random-walk steps depends on the ratio of the likelihoods associated with the random walk's current and proposed new positions.

Starting from an arbitrary initial point X_0 in parameter space, every move is determined by two steps:

1. A proposal Y_n is generated by a random-walk step from position X_{n-1},

$$Y_n = X_{n-1} + SU_n, \tag{3.26}$$

   where $U_n \sim \mathcal{N}(0, \sigma)$. Both the shape matrix S and the width σ of the proposal distribution must be chosen beforehand and kept constant during a run of the algorithm.

2. The proposal is then accepted with the probability

$$\alpha_n := \alpha(X_{n-1}, Y_n) := \min\{1, \pi(Y_n)/\pi(X_{n-1})\}, \tag{3.27}$$

   in which case $X_n = Y_n$, i.e. the walker moves to the proposed position. If the proposal is rejected, then the random walk remains at the current position $X_n = X_{n-1}$.

By recursively following these rules the chain of accepted samples of the algorithm asymptotically converges to the true distribution of π. However, the speed of convergence greatly depends on an optimal choice of both the shape matrix S and the width parameter σ of the proposal distribution. Poor choices lead to abundant rejections (i.e. the chain is stationary most of the time if S is chosen badly or σ is too large) or strong autocorrelations of the samples (i.e., movements are very small if σ is chosen too small, even if S is optimal). Both parameters are however not known in advance and cannot be obtained due to analytical intractability of SWIFT model's likelihood function.

Therefore, we used the *Robust Adaptive Metropolis* (RAM) algorithm by Vihola (2012) which progressively captures the parameters' covariance structure shape and at the same time attains a predefined acceptance rate (see Roberts, Gelman, & Gilks, 1997).

The speed of the adaptation can also be specified parametrically. Although the RAM algorithm is a good strategy for parameter estimation, it is still computationally expensive, as exploration is naturally slow, if subsequent samples are dependent. Furthermore, it is necessary to use several independent chains with randomly dispersed initial values, each requiring a burn-in phase necessary for the sampler to progress to the vicinity of the stationary distribution.

An additional modification of the MCMC algorithm is necessary because of the stochastic pseudo-likelihood function of the SWIFT model. If, by chance, an exceptionally high log-likelihood value is obtained for a proposal, the acceptance rate for the subsequent proposal will be very low, which might stall the chain (Holmes, 2015). Therefore we re-evaluate $\pi(X_{n-1})$ for every iteration of the algorithm, which, however, doubles the computation time of the sampling.

To increase computational efficiency, we introduced parallel computation at two levels. First, while the likelihood of a fixation is dependent on all preceding fixations in the respective fixation sequence, likelihoods of whole fixation sequences can be computed independently from each other and added up later. This procedure enables computing the log-likelihood for independent fixation sequences in F in parallel using a multi-core compute cluster. Second, different chains are independent of each other and can therefore be calculated in parallel as well.

3.3.2 Parameter recovery using simulated data

Before we demonstrate the application of the MCMC framework for the SWIFT model to experimental data, we investigate its performance for simulated data with known parameters. While we tested the likelihood function using single-parameter variation around the true value in Section 3.2.10, we now estimate all four selected parameters (Tab. 3.3) simultaneously using the MCMC procedure for the same dataset. We specified truncated normal distributions centered at parameter ranges (see Tab. 3.3). The standard deviation was set to one half of the estimation range in order to obtain an uninformative prior. We ran 5 independent chains with $N = 4,000$ iterations each and the default adaptation parameter value of $\gamma = 2/3$. The resulting marginal posterior distributions are given in Figure 3.6, where all true parameter values lie within the 40% highest posterior density interval (HPDI). The results suggest that the likelihood-based MCMC framework is very promising for parameter estimation based on data from single participants.

3.3.3 Estimation of parameters based on experimental data

In the next step, we estimated the same parameters for data from an eye tracking experiment. We used the control condition from a larger experimental study on parafoveal processing using the boundary paradigm (see Risse & Seelig, 2019, for a detailed description of the boundary paradigm). We ran 10 chains per participant, each with 4,000

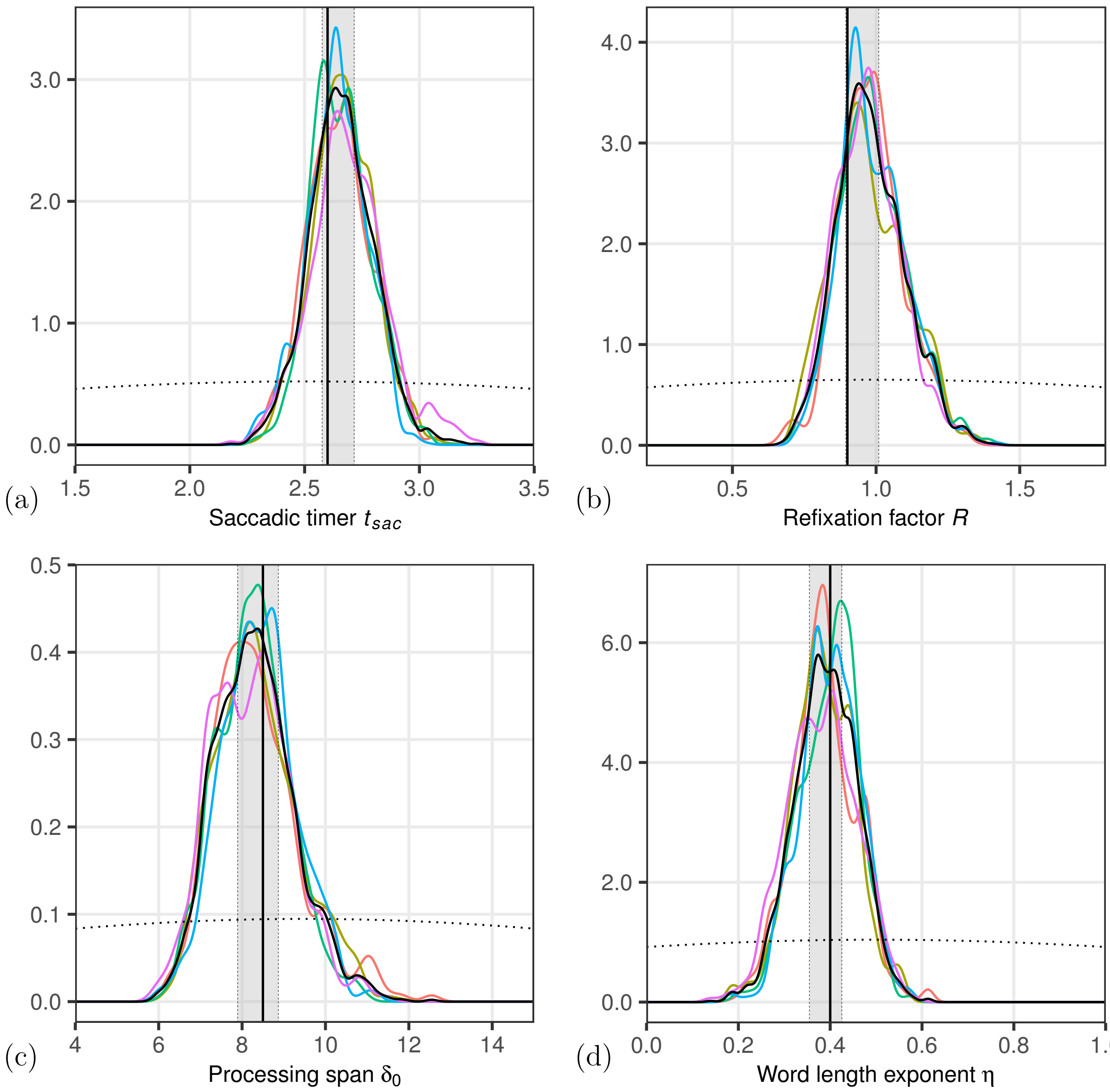

Figure 3.6: Exemplary Posterior distributions of five individual chains (different colors) for four parameters based on simulated data. The black vertical lines indicate the true parameter values. Grey areas indicate the 40% HPDI of all chains. The scale of the parameter range reflects the width of the prior (black, dotted).

iterations. We used the last 2,000 samples (50%) after the burn-in to estimate the posterior density. The resulting marginal posterior densities for a single participant are plotted in Figure 3.8. While there is an increased variance in the posterior densities for the estimation using experimental data compared to the simulated data (Fig. 3.6), we observe clear convergence of the independent chains to a common posterior estimate. Since there is qualitative agreement for the results on simulated and experimental data, the method seems promising to investigate interindividual differences via parameter estimation, which is discussed in the next section.

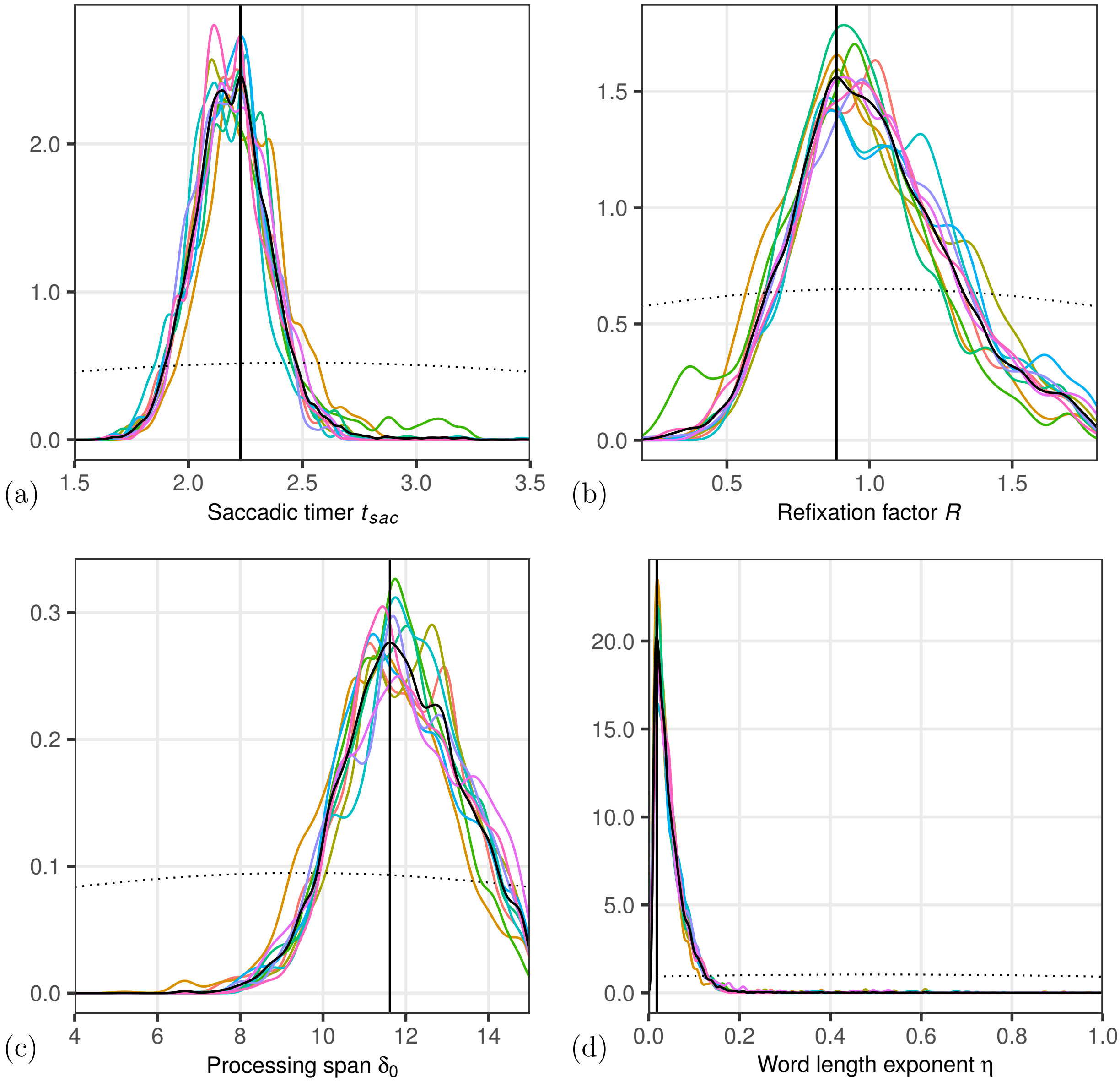

Figure 3.7: Posterior densities for 10 independent chains (coloured) for experimental data from a single participant. The MAP estimator for the pooled chains (black) of each respective parameter is in indicated by the black vertical line. The prior is indicated by the black dotted line.

3.3.4 Interindividual differences and model parameters

In this section we study interindividual differences in model parameters across 34 subjects that served as participants in the experiment by Risse and Seelig (2019). Figure 3.8 shows the posterior densities for all subjects, demonstrating considerable interindividual differences over the model parameters t_{sac}, R, and δ_0, whereas estimates of η fall close to zero.

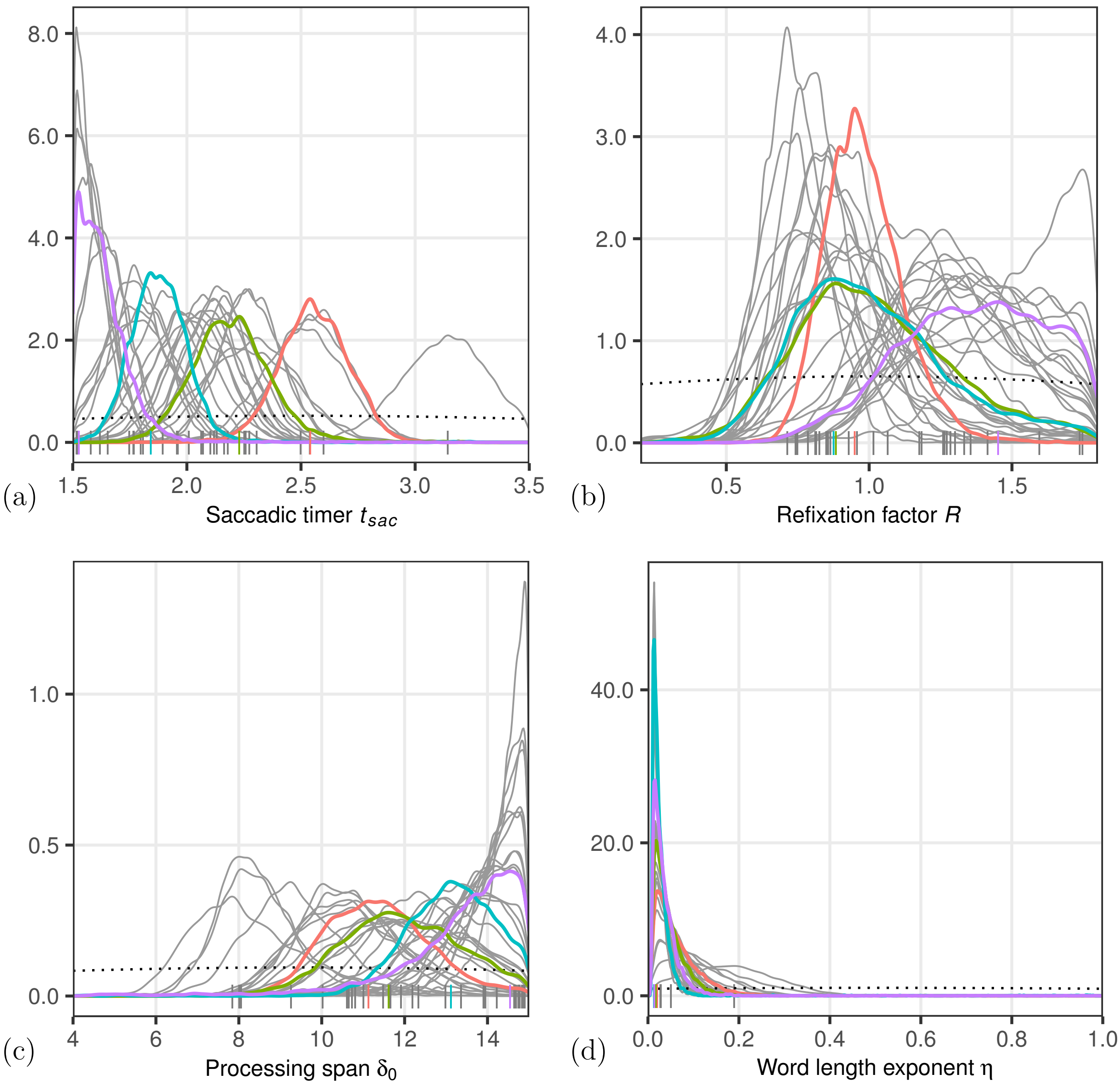

Figure 3.8: Posterior distributions (grey) of 34 participants. Each density is calculated from the pooled data of 10 chains after the burn-in interval. Black ticks at the bottom indicate the MAP estimators for the individual chains. The prior distributions are indicated by the dotted, black line. Curves with the same color correspond to 4 highlighted participants.

A critical question is how much of the differences in reading behavior could be explained by the estimated differences in model parameters. Therefore, we used the maximum a posteriori (MAP) estimator (i.e. the mode) of the pooled chains for each subject as input parameters for the generative model and created a simulated data set that corresponds to the experimental data.

Fixation durations. For both the experimental and the artificial data, we calculated participant-wise averages in different measures of fixation durations. Specifically

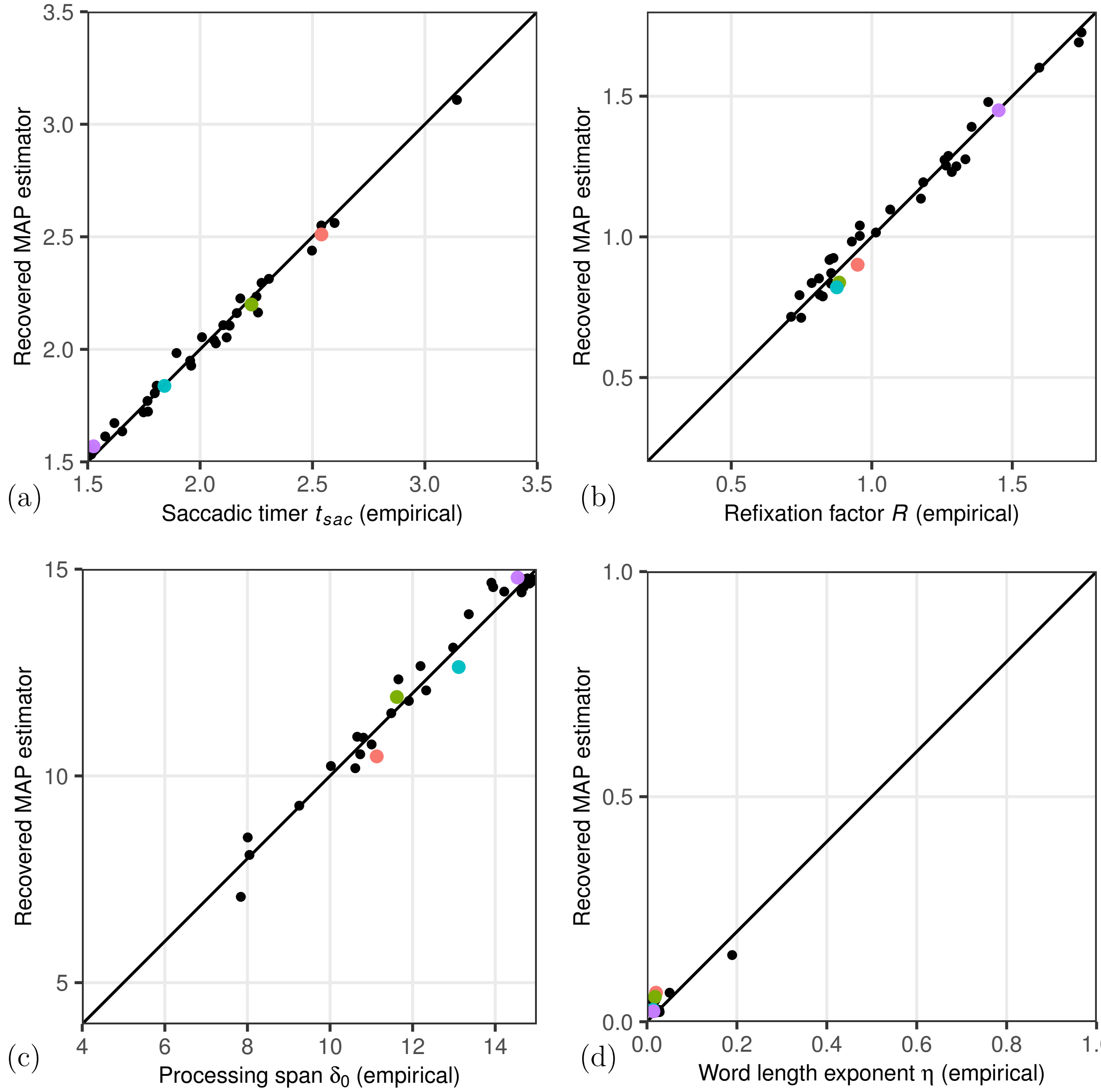

Figure 3.9: Relationship between true parameters (horizontal axis) and estimated parameter values of generated data (vertical axis). Parameters used are the MAP estimators for the experimental data. The coloured points correspond to the same participants as in Fig.3.8.

we compared durations of single fixations (*SFD*; when the word was fixated only once in first-pass), first fixations (*FFD*; when the word was fixated once or more in first-pass), refixations (*RFD*; the second fixation on words, which were fixated more than once consecutively in first-pass), gaze durations (*GD*; the total time spent on a word in first-pass) and total viewing time (*TVT*; the total time spent on a word regardless of first, second or more passes). The results (Fig. 3.10a) indicate a remarkably good fit between the experimental data and model predictions for individual participants for RFD and GD. Mean FFD and SFD generated by the model tend to be slightly underestimated for participants

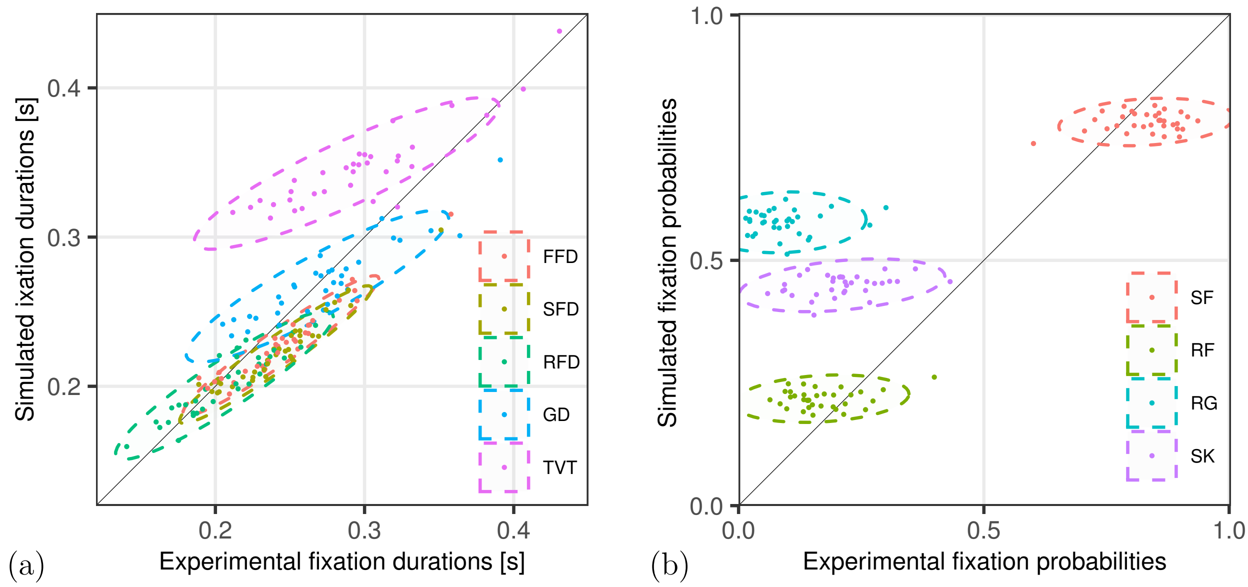

Figure 3.10: (a) Means of different measures of fixation duration for experimental and corresponding simulated data. Each point represents one participant. Simulated data were created using the mean estimated parameters for each respective participant. The coloured ellipses represent the 95% confidence boundaries. (b) Means of word based fixation probabilities. Again each point represents one participant.

with longer initial fixations. Mean TVT, however, is higher in the model predictions than in the experiment. It is important to note that the TVT measure captures more complex gaze behavior, since it also incorporates additional fixation time due to regressions.

Fixation probabilities. Similar to the analysis of fixation durations, we calculated word-based probabilities for single fixations (SF), refixations (RF), regressions (RG), and word skipping (SK) (Fig. 3.10b). While in the experiment words are more likely to receive single fixations as compared to the simulated data, they consequently have a lower probability of receiving refixations. Additionally, the model predicts higher skipping probabilities and also higher probabilities of serving as regression target. It should be noted that the mismatch between experimental and simulated regression probabilities and experimental and simulated TVT (discussed above) is closely related. In general, part of the regressions might be looked upon as a more complicated psycholinguistic measure related to various aspects of post-lexical processing (Rayner, 1998) that cannot be captured in the SWIFT model, while another portion of the regressions might be of oculomotor origin and can be found even in scanning tasks (Nuthmann & Engbert, 2009).

In summary, our results indicate that estimated parameters can explain some of the interindividual differences in fixation durations and fixation probabilities. Thus, the likelihood-based MCMC approach to parameter inference could be applied successfully to estimate model parameters from individual behavioral data.

3.4 Discussion

Current approaches to parameter inference and model comparison (e.g., Reichle et al., 2003) for dynamical cognitive models are insufficient in at least three ways: First, dynamical models need to be tested against time-ordered observations. Second, a likelihood-based procedure is necessary for statistical inference. Third, parameter estimates are needed for individual subjects to explain interindividual differences based on specific model assumptions or components. We set out to solve these three issues in current modeling in computational cognitive science using the SWIFT model of eye-movement control during reading (Engbert et al., 2005) as a case study.

The approach discussed here is fundamentally based on the likelihood function of the model. Therefore, we proposed and investigated the numerical likelihood computation of the SWIFT model. This approach is based on the observation that incremental prediction of fixation positions and fixation durations by the generative model can be exploited to determine the likelihood of the next fixation.

Since the likelihood can be decomposed into a spatial (i.e., fixation position) and a temporal part (i.e., fixation duration), we tried to find separate solutions to both problems. In the spatial part of the likelihood function, internal degrees of freedom (stochastic internal states) could not be integrated out due to numerical efficiency considerations; therefore, we computed a (stochastic) pseudo-likelihood (see Andrieu & Roberts, 2009). In the temporal part, the theoretical likelihood function was unavailable. Therefore, we constructed an approximate likelihood function using a sufficient number of predicted fixation durations from the SWIFT model and KDE for the approximation of the likelihood. In sum, we combined a pseudo-marginal spatial likelihood and an approximated pseudo-likelihood (see Holmes, 2015, for nomenclature) function to obtain the likelihood function of the model (Sisson & Fan, 2011).

Before we applied our framework to real data, we demonstrated that, in a simplified model version with 4 free parameters, we could reconstruct the true parameter values from simulated data. We used a Bayesian approach using MCMC sampling from the posterior distribution based on an adaptive sampling algorithm (Vihola, 2012). The size of the simulated data-set was comparable to a typical experimental data set that is recorded from an individual participant during a one-hour session of eye-tracking experimentation. Next, the same procedure was applied to experimental data. Motivated by the results from simulated data, we estimated model parameters independently for 34 subjects.

Finally, our results indicate that it is possible to relate interindividual differences in reading behavior (characterized by 5 different measures of fixation durations and 4 different measures of fixation probabilities) to differences in the estimated model parameters. Given the typical state-of-the-art models of eye-movement control in reading, this is a major step for generating hypotheses on the observed interindividual differences in a task as complex as reading.

Throughout the current work, we focused on the numerical implementation of the likelihood function for the SWIFT model. Since likelihood-based Bayesian inference turned out to be a viable and sound alternative to ad-hoc parameter estimation procedures, we expect that our approach can be further advanced for both theory building and modeling of interindividual differences. For example, for higher dimensional parameter spaces Differential Evolution MCMC algorithms (see, e.g., Laloy & Vrugt, 2012; ter Braak, 2006; ter Braak & Vrugt, 2008) might be more adequate. Additionally, we expect that a hierarchical Bayesian design will help to increase the stability of the posterior estimates for individual subjects—even if we apply our methods to data sets smaller than used in the current work.

3.5 Acknowledgments

This work was supported by grants from Deutsche Forschungsgemeinschaft (SFB 1294, project B03 to R.E. and S.Re.; SFB 1287, project B03 to R.E. and Shravan Vasishth; grant RI 2504/1-1 to S.Ri.). We acknowledge a grant for computing time from Norddeutscher Verbund für Hoch- und Höchstleistungsrechnen (HLRN, grant bbx00001).

3.6 Appendix A. Experimental data and sentence material

All eye-tracking data used in our simulation studies originate from (Risse & Seelig, 2019), who collected data for an experiment that was a version of the $n+1$ boundary paradigm (Rayner, 1975a) to investigate effects of parafoveal word difficulty on fixation durations and distinguish them from preview benefit effects (see Vasilev and Angele, 2017, for a comprehensive review). Their data is available online at 10.17605/OSF.IO/KZ483.

In the experiment, 34 participants, mostly students of psychology at the University of Potsdam, read 114 single sentences presented on a computer screen while their eyes were being tracked. The simple structured German sentences consisted of six to 12 words with an average length of 9 words. Every sentence contained a gaze contingent invisible boundary before a specific target word. Before the eyes crossed the boundary, the preview of the target word could either be of low, high or medium frequency (i.e. high, low or medium difficulty respectively). During the saccade in which the boundary was crossed, the target word was always exchanged with the medium frequency word. Word frequencies were taken from the dlexDB database (Heister et al., 2011) based on The DWDS corpus: A reference corpus for the German language of the 20th century (Geyken, 2007).

Data treatment and preprocessing. The data were collected using an Eyelink II System (SR Research, Osgoode/Ontario, Canada) with a temporal resolution of 1000 Hz. Since spatial resolution was preprocessed to letter accuracy. Within-letter position was randomized by added small random numbers to avoid artifacts from discretization. Basically, the data used here were treated by the same preprocessing as reported in the statistical analysis of the experiment. Additionally, fixation durations smaller than 25 ms were discarded (550 fixations in 338 trials). Trials that included fixation durations larger than 1000 ms were discarded (45). Trials consisting of less than three fixations were also removed from the data-set. Additionally, re-readings signaled by regressions starting from the second last or last word of the sentence and all subsequent fixations were discarded (5773 fixations). After preprocessing, 30,639 fixations from 3422 trials were included in the data-set for estimation. The implementations of the model, the estimation algorithm, and scripts for analyses and plots, along with the corpus data and fixation sequences are available at 10.17605/OSF.IO/XDKWQ.

3.7 Appendix B. Supplementary data

Supplementary material related to this article can be found online at
https://doi.org/10.1016/j.jmp.2019.102313.

Chapter 4

Predictive modeling of the influence of parafoveal information processing on eye guidance in reading

Stefan A. Seelig[a], Sarah Risse[a], Ralf Engbert[a,b]

[a] Department of Psychology, University of Potsdam, Germany
[b] Research Focus Cognitive Sciences, University of Potsdam, Germany

Running Head: Predictive modeling

submitted Dec. 2020 to *PNAS*.

Abstract

Skilled reading requires information processing of the fixated and the not-yet-fixated words to generate precise control of gaze. Over the last 30 years, experimental research provided evidence that word processing is distributed across the perceptual span, which permits recognition of the fixated (foveal) word as well as preview of parafoveal words to the right of fixation. However, theoretical models have been unable to differentiate the specific influences of foveal and parafoveal information on saccade control. Here we show how parafoveal word difficulty modulates spatial and temporal control of gaze in a computational model to reproduce experimental results. In a fully Bayesian framework, we estimated model parameters for different models of parafoveal processing and carried out large-scale predictive simulations and model comparisons for a gaze-contingent reading experiment. We conclude that mathematical modeling of data from gaze-contingent experiments permits the precise identification of pathways from parafoveal information processing to gaze control, uncovering potential mechanisms underlying the parafoveal contribution to eye-movement control.

4.1 Introduction

High-acuity visual processing is limited to the center of the visual field (the fovea) with an extension of about 2°, which fits a short word at typical font size and stimulus distance. Consequently, humans need to generate fast eye movements (saccades) to move words into the fovea for word recognition during natural reading (Findlay & Gilchrist, 2003). However, the visual field is much larger than that and words are processed, although with lower visual acuity, beyond the fovea (in the parafovea). Here, we report results on the use of word information from the parafovea for eye-movement control during reading. We present an explicit computational model of parafoveal processing in an experimental paradigm. Our approach is fully predictive, i.e., the model is trained under natural reading conditions and makes predictions for the effects of experimental manipulations of the reading process.

A critical concept for information processing during reading is denoted as the perceptual span (McConkie & Rayner, 1975), which is the area of the visual field in which text must be visible for the reader to proceed reading at a normal speed. Experimentally, the perceptual span has been measured by systematically increasing the size of a window of visible text that moves with the readers' gaze across the sentence until readers cease to show significant disruption in their reading behavior (Jordan, McGowan, Kurtev, & Paterson, 2016; McConkie & Rayner, 1975; Rayner et al., 1980). The average size of the perceptual span extends roughly from 3 to 4 letters to the left of fixation to about 14 to 15 letters to the right of fixation and is therefore asymmetric around the fixation location (Rayner, 2009a). Although low-level pre-attentive processes may also contribute to the observed size (i.e., the letter identification span is considerably smaller ranging only up to 7 to 9 letters to the right of fixation) (Underwood & McConkie, 1985), the perceptual span is strongly associated with word recognition processes and thus with the allocation of attention during reading.

4.1.1 Experimental findings on parafoveal processing

The boundary paradigm (McConkie & Rayner, 1975) is among the most frequently used experimental methods to study the effects of parafoveal processing on the timing of the reader's eye movements during reading. Contingent on the reader's gaze position, the parafoveal preview of a target word (word $n + 1$) is manipulated in an *invalid preview* condition (e.g., a random letter string or a different word is presented), while the reader's eyes fixate to the left of it (e.g., before or on the pretarget word n). When a saccade is launched past the location of an invisible boundary placed after the last letter of word n, the preview of word $n + 1$ is changed and replaced by the target word (Fig. 4.1). Readers typically lack awareness of such display changes (Angele & Rayner, 2011; Angele, Slattery, & Rayner, 2016; Matin, 1974; Risse & Kliegl, 2014; White, Rayner, & Liversedge, 2005).

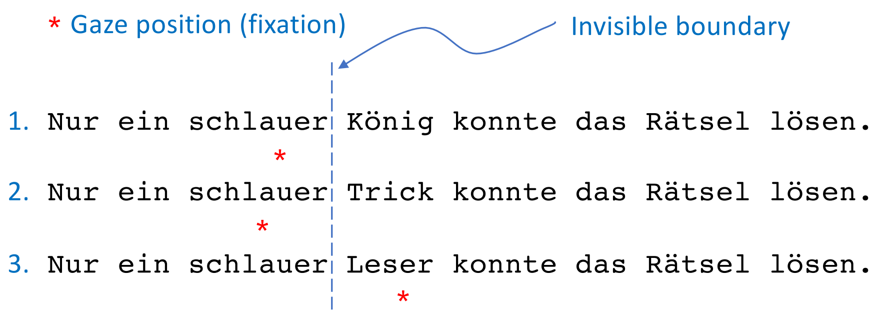

Figure 4.1: The boundary technique as a variant of gaze-contingent displays. The critical word position is to the right of an invisible boundary. If gaze position is to the left of the boundary (*first line*), the preview word is diplayed. The preview is either a high-frequency word (`König`) or a low-frequency word (`Trick`). A saccade crossing the boundary triggers an immediate display change that replaces the preview by a medium-frequency word (`Leser`).

Conversely, in the *valid preview* condition the preview is identical to the target word, thus representing normal reading.

In experiments using the boundary paradigm, readers show differences in fixation durations as a function of the preview condition in which the sentence was presented. The first finding in the boundary paradigm is an effect of the preview validity in fixation durations on the target word $n + 1$ to the right of the boundary. Fixation durations on word $n + 1$ are longer when an invalid preview was presented and shorter when the valid (identical) word was displayed before its fixation (Starr & Rayner, 2001). This difference in fixation durations is typically interpreted as a preview benefit resulting from a headstart of processing the identical preview in parafoveal vision (Schotter et al., 2012). As parafoveal preprocessing reduces the word's remaining processing demand, word recognition times are shorter when the word is finally fixated.

The second finding is an effect of the preview difficulty. Fixation durations are longer when the parafoveal preview was a difficult word (high processing load) and shorter when it was an easy word (low processing load). Manipulating the parafoveal processing load by the preview's lexical frequency (i.e., the frequency of occurrence of a word in a representative text database), preview difficulty effects have not been observed on the pretarget word n before the boundary (Brothers et al., 2017; Risse & Seelig, 2019) but on the target word $n+1$ after the boundary. (Risse & Kliegl, 2012, 2014; Risse & Seelig, 2019; Schotter & Leinenger, 2016) While it seems clear that the preview must have been preprocessed up to its lexical level in parafoveal vision, the precise mechanisms that prolong the critical fixation after the boundary can best be investigated using explicit computational models.

4.1.2 Computational predictions for eye-movement control

Several computational models (Rayner & Reichle, 2010) of eye-movement control have been developed over the last 20 years. Interestingly, model comparisons are limited to qualitative analyses so far (Reichle et al., 2003), mainly due to the lack of adequate statistical methods for model inference of complex process-oriented cognitive models (Schütt et al., 2017). The SWIFT model (Engbert et al., 2005) provides a conceptually favorable architecture in the context of implementing mechanisms for the contributions of foveal and parafoveal processing on eye-movement control; the model provides a platform for studying interactions between foveal and parafoveal processing without major changes of the model principles.

Another prerequisite for the investigation of quantitative predictions is a reliable framework for statistical inference. Recently, we implemented a fully Bayesian framework for parameter inference for the SWIFT model (Rabe et al., 2020; Seelig et al., 2020), which permits parameter identification based on experimental data from single readers in a statistically rigorous way. Therefore, we implement our assumptions on the interaction of fixation duration with foveal and parafoveal processing in the SWIFT model to investigate the potential of various mechanisms in explaining the integration of foveal and parafoveal information during reading.

In the SWIFT model (Engbert et al., 2005) fixation durations are controlled by a random saccade timer that initiates new saccade programs, which accounts for the stochasticity in fixation durations (see Supplementary Note 1). Influences from cognitive word processing are introduced into the model by inhibitory processes. Each word n is represented in SWIFT by an activation $a_n(t)$ at time t under the assumption of parallel processing (Snell & Grainger, 2019). Processing difficulties for low-frequency words produce higher lexical activations on average which delay the start of the upcoming saccade program. Consequently, fixations on difficult words show increased fixation durations (Fig. 4.2). In the latest version of SWIFT (Rabe et al., 2020; Seelig et al., 2020), only the processing of the currently fixated word in the fovea affects the random timer through foveal inhibition. In this study, we investigate additional parafoveal inhibition from activation $a_{n+1}(t)$ to the right of the fixated word n. In two different variants, parafoveal inhibition can act on the timer either immediately ($\tau = 0$) or with temporal delay ($\tau > 0$). The display change was implemented as a reset of the target word's activation values to zero, and restart of processing with the first fixation after the boundary during invalid preview conditions.

For our simulation studies, we adopted a fully predictive framework, where the model was fitted to data of the control condition only (i.e., with valid preview), while data of two invalid preview conditions, (i) invalid high-frequency (HF) preview or (ii) invalid low-frequency (LF) preview, were simulated as quantitative predictions. The difference between the mean fixation durations of the invalid HF and LF preview conditions estimated the preview difficulty effect, whereas the difference between valid and invalid

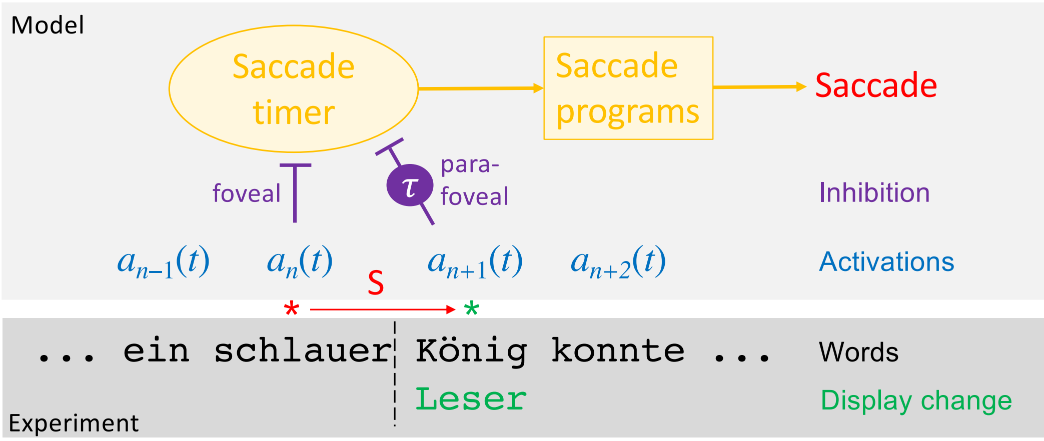

Figure 4.2: Modeling eye-movement control in the boundary paradigm. Experiment: the saccade S triggers the display change from the preview `König` to the fixated word `Leser`. Model: Word-based activations indicate states of lexical processing for each word. The saccade timer that initiates a cascade of processes that produce the saccade. Inhibition of the saccade timer can delay the saccade, which is observed as increased fixation duration. Based on model simulations, we investigated whether inhibition by foveal (word n) and/or parafoveal words (word $n+1$) is more consistent with experimental data. Parafoveal inhibition can be immediate ($\tau = 0$) or delayed ($\tau > 0$).

preview conditions (the latter computed as the average of mean fixation durations in HF and LF conditions, respectively) tested the preview validity effect. The model simulations of the boundary experiment provide a strong test of possible pathways from foveal and parafoveal information to gaze control within a well-defined mathematical model under statistically reliable procedures (Rabe et al., 2020; Schütt et al., 2017; Seelig et al., 2020).

In a first step, we explored to what extent cognitive control mechanisms could suffice to account for the spatio-temporal pattern of preview effects in the boundary paradigm (i.e., no preview difficulty and validity effects on word n but on word $n+1$). Therefore, we extended the model's cognitive control from only foveal (P0) to also parafoveal inhibition. Parafoveal inhibition was either acting immediately (P1) or with a delay of 100 ms (P2) accounting for the slower processing efficiency in parafoveal vision. After estimating model parameters, we determined the mean prediction errors of these three model variants.

In a second step, we further analyzed possible interactions between properties of the experimental method and the oculomotor control system. Therefore, we simulated two different types of saccade cancelation scenarios as response to the display change in the invalid preview conditions, similar to saccadic inhibition (Reingold & Stampe, 2004). The first scenario assumed that a substantial proportion of saccade programs with a probability of $p = 0.5$ is canceled based on the visual disruption when replacing the invalid preview with the target word. The second scenario assumed that the successful cancelation further depends on the stage of preview processing and is more likely when the preview is still in the phase of increasing lexical activations.

4.2 Results

For the boundary experiment (see Methods: Experimental data), we carried out numerical simulations to generate predictions of the SWIFT model with foveal (P0), parafoveal (P1), and time-delayed parafoveal (P2) inhibition. We also investigated three assumptions on possible saccade cancelations due to display changes, i.e., without cancelation (baseline), with saccade cancelation (SC), and with cancelation limited to saccade during the increasing stage of lexical activation (SC-L1). In sum, we investigated nine different models. Since we focus on lexical parafoveal processing in an $n + 1$ boundary paradigm, we restricted our analyses of fixation durations to fixation sequences where a single fixation on word n was followed with a first fixation on word $n + 1$. Sequences with multiple fixations on word n were excluded from analysis in simulated as well as experimental data.

Model parameters were estimated for each participant and model based on the experimental control condition data (see Methods: Bayesian parameter inference). Posterior predictive checks were carried out to ensure successful parameter estimations (see Methods: Evaluation of parameter estimations). Since our approach was to predict the outcome of the experimental boundary manipulations, parameter estimates from the control condition were used for the experimental conditions (see Methods: Implementation of the experimental design).

4.2.1 Fixation durations on the post-boundary word

Simulations of nine model variants generated predictions of the reading behavior in two preview conditions based on parameters fitted to the control condition with valid preview. We evaluated the first fixation duration on the post-boundary word (Fig. 4.3, Table 4.1) after single fixations on word n. Models only incorporating inhibition by foveal processing (P0) predict, on average, the same fixation durations after the boundary when changing from an easy (HF) preview to the target word as compared to when changing from a difficult (LF) preview to the target word. When parafoveal processing difficulties within the processing span were inhibiting the autonomous saccade timer (P1), a difference between HF and LF invalid preview condition emerged and the mean values of the simulated fixation durations became larger when a difficult LF preview was processed in parafoveal vision. The effect of a delay of 100 ms of the inhibiting effect of the parafoveal information (P2) is less obvious and our simulations indicate that the effects of the delay unfold in interaction with saccade cancelations.

The three baseline model variants failed to show a substantial benefit of processing a valid (identical) preview in parafoveal vision as compared to an invalid preview. Only after implementing saccade cancelations (SC) based on the display change in the two invalid preview conditions, the condition means differed and a preview validity effect was observed (see also Figure 4.4b). Saccade cancelation further seemed to interact with the parafoveal inhibition mechanisms. In the saccade cancelation models (SC), the mean

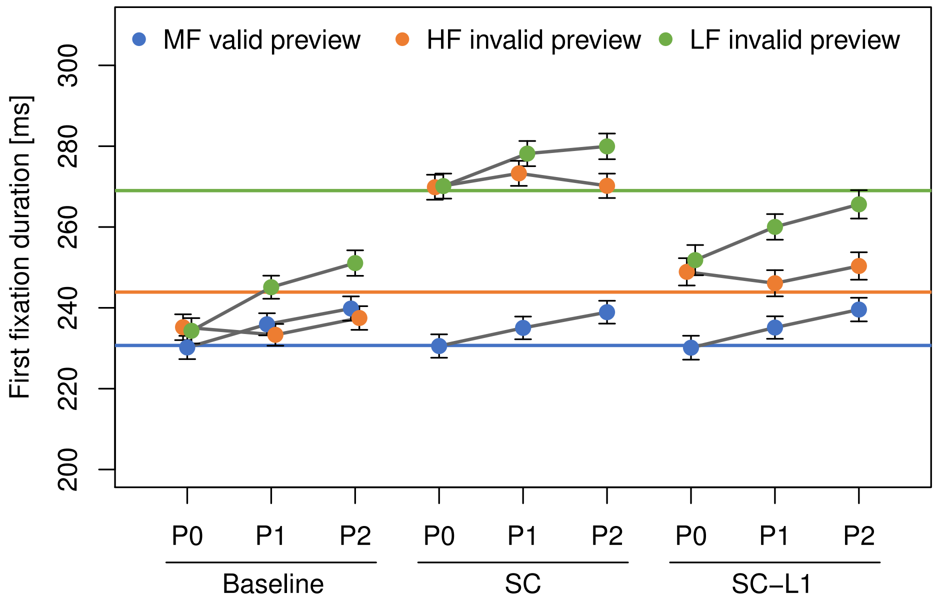

Figure 4.3: Model comparison of $n+1$ fixation durations as a function of three preview conditions. Horizontal lines reflect the empirical condition means using the same color legend. P0: Foveal inhibition only. P1: Foveal and parafoveal inhibition. P2: Foveal and delayed parafoveal inhibition. Baseline: Simple processing reset after display change. SC: Additional saccade cancelation. SC-L1: Saccade cancelation during lexical processing (L1) only.

difference between HF and LF invalid fixation durations in the presence of immediate parafoveal inhibition (P0) was smaller than if parafoveal inhibtion was delayed (P1). In the baseline models (without saccade cancelation) and the processing-dependent saccade-cancelation models (SC-L1), however, this difference was almost of the same size and did not differ much between the two parafoveal inhibition variants.

4.2.2 Predicted preview effects

Figure 4.4 and table 4.1 summarize the results of the nine model variants with respect to the spatio-temporal pattern of parafoveal preview effects in the boundary paradigm. We can conclude that parafoveal inhibition is crucial for the model to account for the novel preview difficulty effect in $n+1$ fixations after the boundary (P1 and P2 models). However, to also account for the classical preview validity effect, the model needs additional mechanisms. Implementing this as a cognitive effect via inhibition on the labile saccade program succeeds in showing a reliable preview validity effect but on the cost of the preview difficulty effect. Thus, the preview validity effect is best explained by saccade cancelations after a display change with only a small portion of cognitive preprocessing benefits (i.e., 4 to 6 ms reflected in the preview validity effect of the baseline models, Figure 4.4b). Moreover, the cognitive preview benefit portion is already fully developed in the P0 baseline model (i.e., no further increase in the P1 and P2 baseline models).

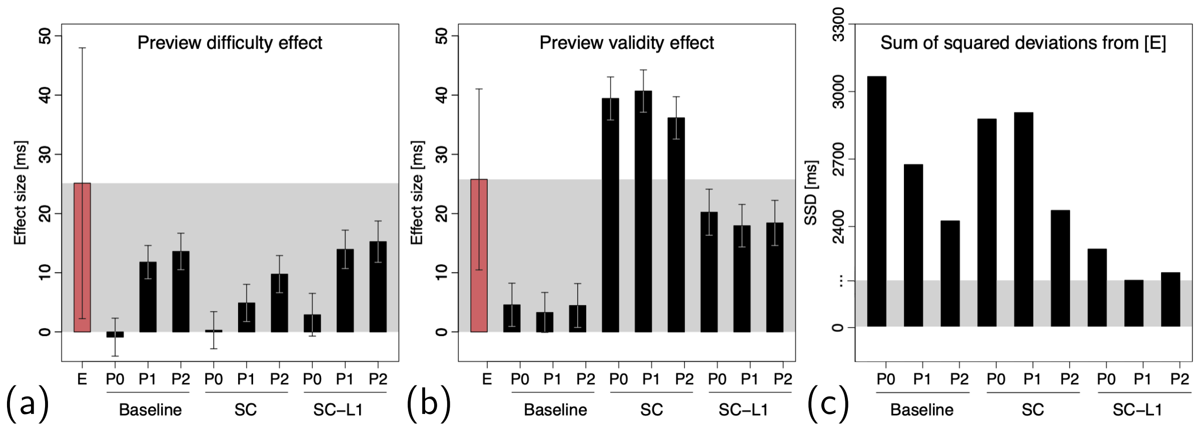

Figure 4.4: Predicted parafoveal preview effects for the first fixation after the boundary. (a) Preview difficulty effect. Smallest difference to experimental effects is observed for baseline models P1/P2 and SC-L1 model P1/P2. (b) Preview validity effect. Smallest difference from experimental effects obtained for SC-L1 models. (c) Overall model scoring. The mean sums of squared deviations from experimental effect sizes indicate best performances for models that combine parafoveal inhibition with processing dependent saccade cancelation.

Thus, in contrast to the preview difficulty effect, the preview validity effect is fully accounted for by foveal inhibition and does not require further parafoveal inhibition. At the qualitative level, the models P1 and P2 with processing-dependent saccade cancelations (SC-L1) show both the best account of preview difficulty and preview validity effects in fixation $n+1$ after the boundary (Fig. 4.4c).

Table 4.1: Mean first fixation durations on word $n+1$ and preview effect sizes in the experiment and model simulations, with 95% confidence intervals in parentheses. CIs are smaller in simulations due to the increased data volume. The models were scored using the mean sum of squared deviations (mSSD) of simulated preview effects from experimental results per participant.

		Baseline			SC			SC-L1		
	E	P0	P1	P2	P0	P1	P2	P0	P1	P2
First fixation durations by condition										
HF	243.9 (13.6)	235.2 (3.2)	233.3 (2.7)	237.5 (2.9)	269.8 (3.1)	273.3 (3.1)	270.2 (3.0)	248.9 (3.4)	246.1 (3.2)	250.4 (3.4)
MF	230.7 (10.0)	230.2 (2.9)	235.9 (2.7)	239.8 (3.0)	230.6 (2.9)	235.0 (2.8)	238.9 (2.8)	230.1 (3.0)	235.1 (2.8)	239.6 (2.9)
LF	269.0 (18.0)	234.3 (3.1)	245.1 (2.9)	251.1 (3.2)	270.1 (3.1)	278.2 (3.1)	280.0 (3.2)	251.8 (3.7)	260.0 (3.2)	265.6 (3.5)
Preview effects										
Difficulty	25.1 (22.9)	−0.9 (3.2)	11.8 (3.2)	13.6 (3.2)	0.3 (3.1)	4.9 (3.1)	9.8 (3.1)	2.9 (3.6)	13.9 (3.6)	15.2 (3.6)
Validity	25.8 (15.3)	4.6 (3.7)	3.3 (3.7)	4.5 (3.7)	39.4 (3.6)	40.7 (3.6)	36.1 (3.6)	20.2 (3.9)	17.9 (3.9)	18.4 (3.9)
mSSD	-	3066.4	2675.3	2424.9	2877.7	2906.5	2471.2	2299.7	2160.6	2194.9

4.3 Discussion

In the current study we investigated mechanisms for a dynamical model of eye-movement control during reading that can account for preview effects which appear in two flavors, the preview validity and preview difficulty effect. We used data from an experiment (Risse & Seelig, 2019) with an $n + 1$ boundary paradigm and three different preview difficulties to estimate posterior distributions of parameters in the SWIFT model of eye movements during reading (Engbert et al., 2005). Parameters were estimated independently for three different implementations of cognitive influences on fixation durations: Inhibition of saccade programming by foveal processing only (P0), additional inhibition by parafoveal processing (P1), and delayed parafoveal inhibition (P2). Estimations were restricted to data from the valid preview condition. Based on the obtained posterior distributions over the model parameters, we predicted fixation sequences for all experimental conditions in the estimated model variants. Each model variant was further crossed with three implementations of the effects of the display change occurring in boundary paradigms. In the baseline models, word processing was simply restarted after the display change, whereas in the two saccade cancelation models the display change could also impede saccade programming either generally, or coupled to the lexical stage of word processing. From the simulated data we calculated the effect sizes for the preview validity effect and the preview difficulty effect for the first fixation after the boundary.

The baseline model P0 with only foveal inhibition did not yield preview difficulty effects, but it was sufficient to elicit a small effect of preview validity. The effect is brought about by the reset of activation of the target word at the onset of its first fixation during invalid preview conditions. While during valid conditions processing of the target word is already in an advanced stage and completed soon after the fixation onset, during invalid conditions processing must start over, resulting in a longer period during which foveal inhibition can influence the random saccade timer. The reason why this influence is small, however, lies in the delay between the onset of saccade programming and the onset of the saccadic movement. Once the random saccade timer has initiated a saccade program that later elicits a gaze shift, foveal inhibition can no longer affect the current fixation duration, even if the fixation still lasts for considerable time.

The introduction of parafoveal inhibition in model P1 did bring about a preview difficulty effect through the mechanism described above. When the model fixates word n and has initiated a saccade program to word $n + 1$, inhibition resulting from the parafoveal preview affects the duration of the upcoming fixation, therefore increasing durations of the upcoming fixation on the target word $n + 1$ in case of LF previews, as compared to HF previews. The impact of parafoveal difficulty was strong enough for MF previews to produce longer subsequent fixation durations than HF previews in the baseline model, despite the difference in the validity condition. Delaying the influence of parafoveal information by 100 ms in the P2 models shifts the evolution of activation more consistently into the time window where the saccade timer's activity is related to the fixation duration on the target word. This affects fixation durations in LF conditions more than in HF conditions, which likely is the result from a dynamical interaction of word frequency with the increase in fixation duration itself.

Display change induced saccade cancelation was introduced as a mechanism to reflect the disruption of visual stimulus continuity that occurs in the case of a display change. Here, if a labile saccade stage is active during the display change, it is aborted with a fixed probability of 50%. A new labile stage can then be initiated in the regular way by the main saccade timer. Unlike with regular saccade cancelation, where an ongoing labile stage is canceled and immediately replaced by a new one (initiated by the main timer), in display change induced saccade cancelation the labile stage is aborted without an immediate replacement. This substantially increases some fixation durations after the display change (i.e., 50% of them) and induces a strong validity effect (see Fig. 4.4).

The third model type was used to investigate processing-dependent saccade cancelation. Psycholinguistic theory suggests that word processing can often be approximated by a two stage process and consists of lexical and post-lexical stages. Research indicates that the visually presented word stimulus is more important during lexical processing (Schotter & Leinenger, 2016) (whereas post-lexical integration can proceed even when the visual representation is absent), which should be reflected in the model. Hence, the display change sensitive saccade cancelation was coupled to the word processing stages within SWIFT, where the epoch of rising activation represents the lexical processing (L1), and the epoch of falling activation represents post-lexical processing (Engbert et al., 2005) (L2). In these model variants (going by SC-L1) labile stages can only be canceled if word processing was still in the earlier L1 stage (although activation reset was done for every invalid preview on crossing the boundary, irrespective of the stage of word processing). This reduced the size of the effect of saccade cancelation on the mean fixation duration in invalid preview conditions, resulting in a pattern more aligned with the data observed in the experiment.

In our simulation study, we used the SWIFT model as a platform for the different variants of parafoveal processing. The parallel processing framework is an open architecture for testing effects of distributed processing (Snell & Grainger, 2019). The recently published OB1-Reader model (Snell, van Leipsig, Grainger, & Meeter, 2018) proposed how letter-level visual and lexical processing could be successfully integrated into a model of eye-movement control. Based on such extensions, we would expect even more specific predictions on the boundary paradigm and its variants.

Finally, it is important to stress that the results presented in this work heavily rely on the success and quality of the parameter estimation. Parameter inference based on individual readers' experimental data might be a breakthrough for process-oriented modeling (Rabe et al., 2020; Schütt et al., 2017; Seelig et al., 2020), since interindividual differences are often comparable in size to the observed effects. Here we exploited the full potential of interindividual differences by running predictive simulations separately for each participants.

4.4 Methods

4.4.1 Experimental data

Experimental data were collected in an experiment of single sentence reading with 34 participants (Risse & Seelig, 2019). Each participant read 114 sentences on a computer monitor in a single session, while their eyes were being tracked. The experiment used the gaze contingent boundary paradigm, where an invisible boundary is placed between two adjacent words. In the beginning of a trial the word displayed to the right of the boundary corresponded to one of three different preview conditions. Those preview words had similar grammatical function and word lengths but could either be of high frequency (HF), medium frequency (MF) or low frequency (LF). Then, as soon as the eyes first crossed the boundary towards the preview, the preview was replaced by a target word of medium frequency. In the MF condition the preview and the target word were identical. The process of replacing the word on the monitor was implemented to be quick enough, that the saccadic movement which had triggered the boundary would envelope the display change event. Of the 3,521 fixation sequences in the collected data, only fixation sequences in the MF condition were selected for parameter estimation. Sequences containing less than three fixations or fixations longer than one second were not considered in the estimation. Additionally, all fixations after regressions from the last or second to last word were removed. This left 1,139 fixation sequences with a total number of 10,172 fixations from 34 participants.

4.4.2 Bayesian parameter inference

The parameter estimations were conducted using a Python implementation (Shockley, Vrugt, & Lopez, 2017) of the DREAM$_{ZS}$ algorithm (Laloy & Vrugt, 2012) from the class of Metropolis-Hastings Markov chain Monte Carlo (MH-MCMC) algorithms (Hastings, 1970). In a Bayesian framework MH-MCMC algorithms use a random walk strategy to iteratively build up a sampling distribution which eventually converges to the posterior distribution $P(\theta|y)$ of parameters θ given the data y. Starting with the chains randomly dispersed in parameter space Θ, the sampler generates new proposals from perturbations of the latest positions of the chains at each iteration. The proposals are integrated into the chain depending on their acceptance probability.

For each of the three models, parameter estimations were conducted using three chains per participant, with $20,000$ iterations per chain. As priors for the parameters, we used Gaussian distributions truncated at one standard deviation, with ranges according to Table 4.2. Since calculating the likelihood in SWIFT uses simulations and approximations, whereby the likelihood is inflated with a stochastic error (Seelig et al., 2020), the DREAM$_{ZS}$ algorithm had to be slightly modified. A stochastic likelihood can have adverse effects on the algorithm's rate of convergence. Originally, at any given position of a chain the likelihood is evaluated only once (Laloy & Vrugt, 2012). However, stochastic fluctuations of likelihood values can impede the calculation of the acceptance ratio and introduce long periods of stagnation in the evolution of chains where no proposals are accepted. To circumvent this, the modified algorithm newly evaluates the likelihood for the latest chain position at every iteration. While this doubles the computational costs, it also prevents the algorithm from becoming stuck. One ramification of this approach is the notion, that the sampling distribution is no longer converging to the posterior distribution, but rather some form of convolution with the error distribution imposed by the likelihood approximation (Seelig et al., 2020).

Numerical calculations were carried out on the high-performance computing cluster of the Norddeutscher Verbund für Hoch- und Höchstleistungsrechnen (HLRN). Parallel computation was used at the level of the estimation algorithm, as well as the level of trials within each participant-wise model evaluation, respectively.

4.4.3 Implementation of the experimental paradigm

The experimental manipulation involved a display change event where, during invalid conditions, an invalid word preview is replaced with a target word. In simulations of the valid MF condition, no changes in model architecture had to be made, since this condition represented normal reading. In the HF and LF conditions the word frequency of the word $n+1$ corresponded to the respective invalid previews by the beginning of a trial. Once the model had finished the execution of a saccade past the last letter of word n, the frequency of word $n+1$ was changed to represent the MF word. Additionally, the activation values of the target word were reset to zero and the processing stage was reset to the first stage.

Parafoveal inhibition (P1) was implemented in a similar fashion as the existing mechanism for foveal inhibition (Seelig et al., 2020). The numerical values of the word activations were multiplied with their respective inhibition factors (see Table 4.3), before modulating the transition rate of the saccade timer. The delay was implemented using a memory array, where activations of parafoveal words were retained together with their time signature, so they could be recalled after the delay of 100 ms.

Two variants of the model implement saccade cancelation in the invalid preview conditions as a result of the display change. In SWIFT, the two-stage process of saccade programming can only be aborted during the first, labile stage, but not during the non-labile, second stage. In the event of a successful cancelation, a new saccade program is immediately initiated, starting with a labile stage, thereby increasing the duration of the current fixation. The causes of cancelations are now extended to include display change events. When a labile stage is active during the time of the display change in the SC model, it is canceled with a probability of $p = 0.5$. In the SC-L1 model it is also required that the second processing stage of the preview word $n+1$ has not yet been reached (this is checked before any variables are reset as a result of the display change).

Artificial fixation sequences were simulated for all subjects with parameter combinations specific to the subjects' estimation results. For each sentence a different set of parameters was randomly sampled from the posterior distribution of the respective participant (Rabe et al., 2020), and per sentence 10 sequences were generated. For the analysis of preview effects, the simulated sentences were processed in the same way, as the experimental data. To keep results comparable, parameters were sampled from the posterior distributions once per participant and sentence. This set of parameters was then used in all simulations for the respective participant in the Baseline, SC, and SC-L1 models and all conditions for the same sentences. Simulations of three participants were excluded from all further analyses due to computer error.

4.4.4 Evaluation of parameter estimations

For a first analysis of the effects of the specific model implementation on the parameter estimation, based on the posterior distributions we calculated the estimation mean of the subjects median parameter values, to compare them with the lower and higher margins of 30% highest posterior density intervals (HPDIs) of the posteriors pooled over participants (Table 4.2). We observe that variability between estimations is lowest for parameters related to spatial aspects of oculomotor control, and higher for parameters concerned with temporal control of saccade timing and word processing.

Posterior predictive checks (Schad, Betancourt, & Vasishth, 2020) were done for the three sets of simulations P0, P1 and P2. Each set consists of 31 distinct simulations based on the posterior distributions of individual participants. The posterior distributions correspond to the experimental control condition (MF). We calculated a set of common summary statistics (Table 4.4, Figure 4.6) from the data of all experimental conditions and the simulated data, in order to cross validate the model fit. Significant Pearson correlations coefficients indicate good agreement of experimental and simulated data in the HF and LF conditions, which were not used in the parameter estimations.

4.5 Data Availability

The experimental data used in this study were published before (Risse & Seelig, 2019) and were made publicly available via the Open Science Framework (DOI `10.17605/OSF.IO/KZ483`). Simulated data are accessible with the source code of the model (see below).

4.6 Code Availability

Source code used for numerical simulations, analyses, and plotting as well as other project-related files are made available via the Open Science Framework (DOI `10.17605/OSF.IO/gdsn7/`).

4.7 Competing Interest Statement

The authors declare no competing interests.

4.8 Acknowledgements

This work was supported by grants from Deutsche Forschungsgemeinschaft (CRC 1294, project no. 318763901 to R.E. and RI 2504/1-1 to S.R.). We acknowledge support for computing time provided by Norddeutscher Verbund für Hoch- und Höchstleistungsrechnen (HLRN, grant bbx00001).

4.9 Author contributions statement

All authors conceived the research, S.A.S. carried out the numerical simulations. S.A.S. and S.R. analyzed the data. All authors wrote and reviewed the manuscript.

4.10 Supplementary Materials

This file includes:

- Supplementary Note 1 (Mathematical details of the SWIFT model)
- Figure 4.5 (Posterior densities from simulations)
- Figure 4.6 (Scatter plot of simulated and experimental measures)
- Table 4.2 (Results from parameter estimation)
- Table 4.3 (Inhibition parameters in model variants)
- Table 4.4 (Correlations between simulated and experimental measures)

4.10.1 Supplementary Note 1 (Mathematical details of the SWIFT model)

In the SWIFT model (Engbert et al., 2005) we proposed two independent mechanisms for target selection and saccade timing, coupling via word-based activations. Word activations represent the current state of word processing over time. At the same time, word activations control target selection for an upcoming saccade and modulate fixation durations via delay of upcoming saccades. The internal state of the model (Seelig et al., 2020) at time t can be written as $n = (n_1, n_2, ..., n_{4+N_W})$ with $n_1, ..., n_4$ representing saccade timers and $n_5, ...n_{4+N_W}$ word activations, where the total number of words in a given sentence is denoted by N_W. Word activations increase during lexical processing and decrease during postlexical processing. All random variables n_i are discrete, so SWIFT is a continuous-time, discrete state random walk model (which can be simulated efficiently via its master equation (Seelig et al., 2020)).

Within a processing span centered at the current gaze location, words are processed in parallel (Snell & Grainger, 2019). The eccentricity of letter j in word i is given by $\epsilon_{ij}(t)$, which is time-dependent due to changes of gaze position via saccades. The spatial extension of the processing span with δ letter spaces to the left and to the right of fixation is a very important parameter. We assume an inverse parabolic processing function, which is the dependence of the processing rate $\lambda(\epsilon)$ from eccentricity ϵ, i.e.,

$$\lambda(\epsilon) = \lambda_0 \cdot \begin{cases} 1 - \epsilon^2/\delta^2\,, & \text{for} \qquad |\delta| \leq \epsilon \\ 0\,, & \text{otherwise} \end{cases} \,, \tag{4.1}$$

where $\lambda_0 = 3/4\delta$ is a normalization constant. The processing rate $\Lambda_i(t)$ for word i at time t is given as

$$\Lambda_i(t) = L_i^{-\eta} \sum_{j=1}^{L_i} \lambda(\epsilon_{ij}(t)) \,, \tag{4.2}$$

with word length L_i and parameter η as an exponent determining the influence of word-length.

A word's activation increases with rate $\Lambda_i(t)$ during lexical processing. When the word-frequency dependent maximum is reached, post-lexical processing starts with a decrease in activation determined by the same processing rate. Additionally, there is a decay rate ω accounting for memory leakage effect (Rabe et al., 2020). Saccade target selection is a stochastic process with targeting probability $\pi(m, t)$ for word m at time t controlled by relative activation, i.e.,

$$\pi(m, t) = \frac{a_m(t)}{\sum_{j=1}^{N_W} a_j(t)} \,. \tag{4.3}$$

Finally, saccades are generated with random inter-saccade intervals (Seelig et al., 2020). To account for word difficulty effects, random timing is inhibited by foveal activation. Thus, the rise-rate of the saccade timer is modulated by a factor $(1 + h\, a_k(t))^{-1}$, where $a_k(t)$ is the foveal (fixated) word k, so that high activation delays an upcoming saccade and, therefore, prolongs ongoing fixation duration. In Figure 4.2, we illustrate how the concept of foveal inhibition is generalized to investigate influences of parafoveal processing; in this case inhibition acts by slowing factor $(1 + h\, a_{k+1}(t))^{-1}$ from word $k+1$. Correspondingly, delayed parafoveal inhibition is given by a factor $(1 + h\, a_{k+1}(t - \tau))^{-1}$.

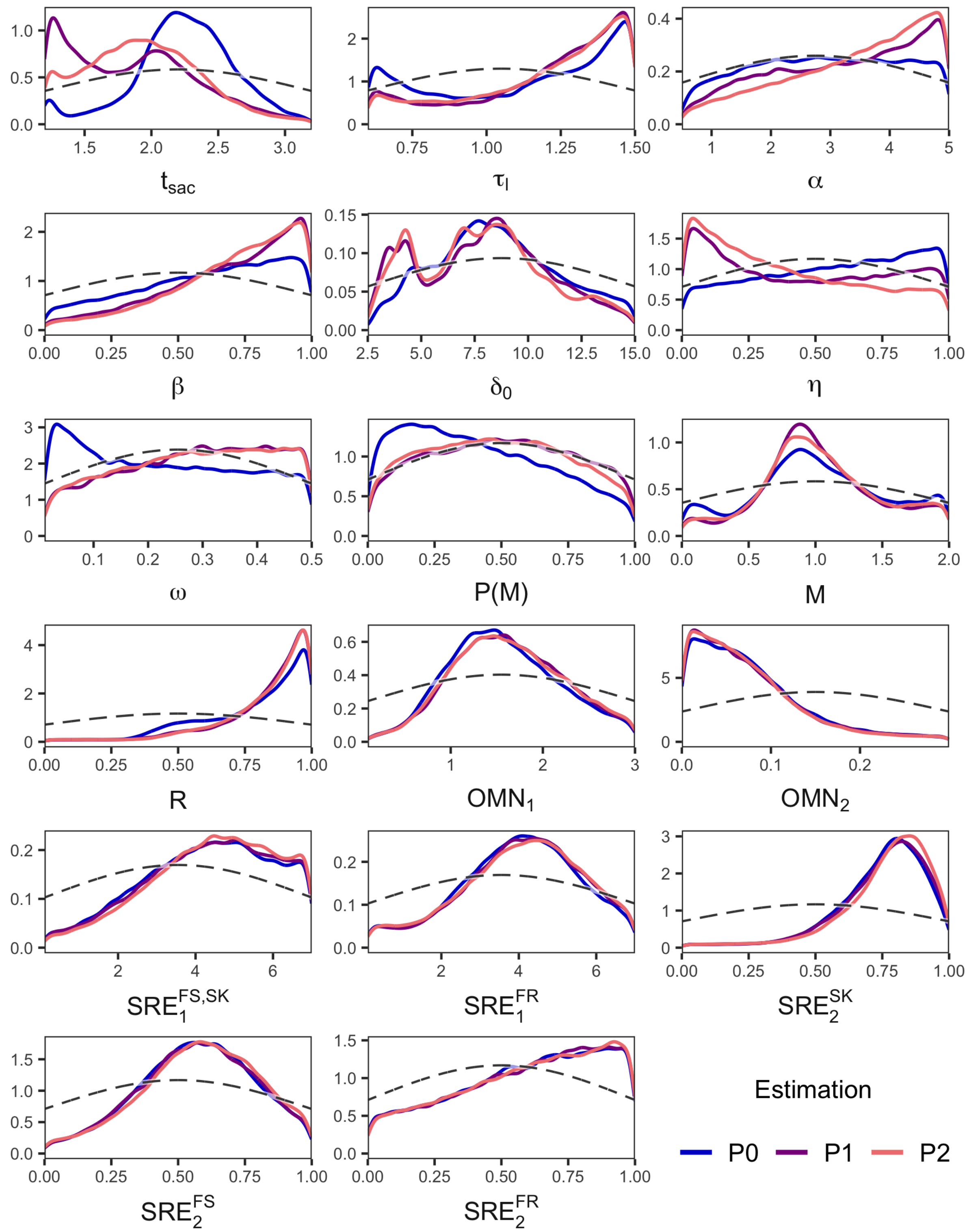

Figure 4.5: Posterior densities for all estimated parameters averaged over participants. Priors (truncated Gaussians) are given by dashed lines. The posteriors for the different model variants are indicated by different colors.

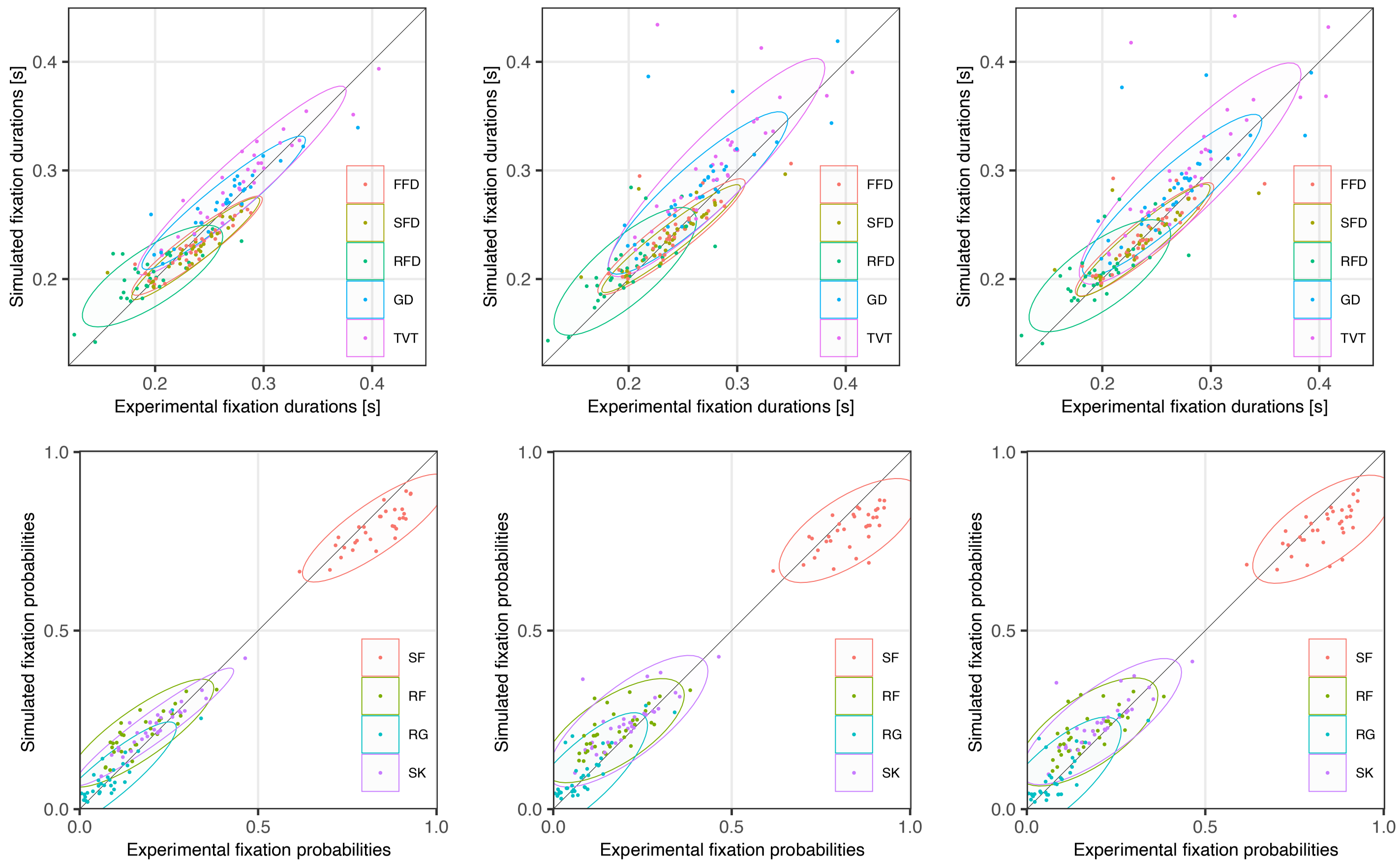

Figure 4.6: Correlation between common summary statistics for experimental and simulated data in the P0 (right column), P1 (middle column) and P2 (left column) dataset without saccade cancelation (baseline model; right). In all plots each point represents the means of experimental and simulated measures for a single participant.

Table 4.2: Parameters estimated for the three model variants. The range values indicate the boundaries of the Gaussian prior distributions which were symmetrically truncated at ±1 standard deviation.

		Range		P0			P1			P2		
Parameter	Symbol	lower	upper	min	HPDI (sd)	max	min	HPDI (sd)	max	min	HPDI (sd)	max
Saccade timer	τ_{sac}	120	320	213	224 (30)	239	159	187 (35)	206	175	195 (30)	210
Processing span	δ_0	2.5	15.0	7.28	8.35 (2.32)	9.52	6.79	7.7 (2.48)	9.15	6.68	7.68 (2.54)	8.96
Decay of activation	ω	0.01	0.5	0.15	0.22 (0.05)	0.3	0.23	0.29 (0.04)	0.36	0.22	0.29 (0.03)	0.35
Lexical difficulty: Slope	β	0.0	1.0	0.5	0.63 (0.1)	0.75	0.62	0.71 (0.12)	0.83	0.64	0.73 (0.1)	0.83
Lexical difficulty: Intercept	α	0.5	5.0	2.3	2.87 (0.61)	3.53	2.69	3.26 (0.67)	3.97	3.05	3.51 (0.62)	4.13
Word length exponent	η	0.0	1.0	0.43	0.58 (0.13)	0.72	0.26	0.46 (0.19)	0.62	0.22	0.37 (0.15)	0.5
Mislocated fixation probability	$P(M)$	0.0	1.0	0.27	0.41 (0.15)	0.52	0.37	0.51 (0.15)	0.63	0.35	0.49 (0.14)	0.61
Mislocated fixation factor	M	0.0	2.0	0.83	1.01 (0.32)	1.2	0.85	1.01 (0.25)	1.14	0.85	1.02 (0.26)	1.16
Labile saccade program	τ_{lab}	60	150	1.02	1.14 (0.24)	1.32	1.16	1.22 (0.2)	1.35	1.14	1.21 (0.19)	1.35
Refixation factor	R	0.0	1.0	0.75	0.79 (0.15)	0.91	0.82	0.84 (0.13)	0.93	0.81	0.84 (0.12)	0.93
Oculomotor noise: Intercept	omn_1	0.1	3.0	1.3	1.56 (0.36)	1.78	1.36	1.61 (0.34)	1.85	1.36	1.61 (0.34)	1.85
Oculomotor noise: Slope	omn_2	0.0	0.3	0.04	0.07 (0.04)	0.09	0.04	0.07 (0.04)	0.09	0.04	0.07 (0.04)	0.08
SRE: Intercept refixations	sre_1^{RF}	0.1	7.0	3.46	3.99 (1.04)	4.65	3.56	4.08 (1.03)	4.76	3.54	4.07 (1.05)	4.78
SRE: Slope refixations	sre_2^{RF}	0.0	1.0	0.49	0.61 (0.14)	0.74	0.49	0.62 (0.13)	0.75	0.5	0.61 (0.13)	0.75
SRE: Intercept forw. and skip.	$sre_1^{FS,SK}$	0.1	7.0	3.66	4.34 (1.12)	5.09	3.74	4.39 (1.06)	5.15	3.86	4.5 (1)	5.23
SRE: Slope forward saccades	sre_2^{FS}	0.0	1.0	0.49	0.57 (0.12)	0.66	0.49	0.57 (0.12)	0.66	0.5	0.58 (0.12)	0.68
SRE: Slope skipping saccades	sre_2^{SK}	0.0	1.0	0.71	0.74 (0.12)	0.82	0.72	0.75 (0.12)	0.83	0.74	0.77 (0.12)	0.85

Table 4.3: Model variants. Fixed parameter values were used for the three different implementations of foveal and parafoveal inhibition.

	Inhibition parameter	
Model variant	foveal	parafoveal
No parafoveal inhibition (P0)	2.0	0.0
Parafoveal inhibition (P1)	2.0	3.0
Delayed parafoveal inhibition (P2)	2.0	3.0

Table 4.4: Pearson correlation coefficients for correlations between experimental and simulated mean fixation durations and fixation probabilities across participants for all estimations and conditions.

Summary statistics	P0			P1			P2		
	HF	MF	LF	HF	MF	LF	HF	MF	LF
Fixation durations									
Single Fixation Duration	.67	.71	.72	.78	.84	.77	.74	.80	.76
First Fixation Duration	.81	.81	.76	.70	.74	.67	.69	.72	.66
Re-Fixation Duration	.49*	.74	.49*	.50*	.74	.56	.52	.71	.55
Gaze Duration	.68	.72	.73	.77	.81	.76	.74	.78	.77
Total Viewing Time	.78	.81	.76	.72	.75	.73	.68	.73	.70
Fixation probabilities									
Single Fixation	.73	.76	.74	.73	.76	.75	.74	.78	.74
Refixation	.71	.74	.68	.69	.68	.68	.67	.68	.66
Regression	.79	.74	.76	.81	.80	.80	.76	.71	.75
Skipping	.71	.74	.68	.69	.68	.68	.67	.68	.66

* $p < .01$; all other values $p < .001$

Chapter 5

General discussion

The question of the extent to which words are processed before they are fixated during natural sentence reading is part of a larger complex of controversies on the relationship of cognition and eye movements. The debate plays a role in motivating the conception of several generative computational models which implement different aspects of the conflicting positions, with two specific models dominating the debate. While all models are capable of reproducing certain key features of reading behavior, it is generally difficult to compare their performance. This is in part due to the arbitrariness of their parameterization.

In the present work we confirm that parafoveal processing extends to the lexical level, which has previously been rejected. As the mechanism behind the finding remains unclear, we compare different implementations in the framework of the SWIFT model. One requirement for such quantitative comparisons is the identification of model parameters, to counteract the effects of altering the model structure. This is achieved in a Bayesian framework using Markov Chain Monte Carlo techniques with an approximative Likelihood function for the model. In the coming sections I summarize the results of the three studies and discuss the importance of lexical processing in the context of the debate on parallel vs. serial word processing, and, finally, the role of model comparisons for further research.

5.1 The present studies

In Chapter 2 we investigated the effects of preview difficulty in an $n + 1$ boundary paradigm. Words of low, medium, and high word frequency served as previews, whereas the target word was always the medium frequency word. This allowed for a clear distinction between effects of preview validity and preview difficulty. Comparisons between valid and invalid preview conditions showed shorter fixation durations on word $n + 1$ for valid previews, which is a common finding. Further, in comparisons between the invalid preview conditions we found an effect of word frequency on fixation durations for fixations on the target word $n + 1$. Specifically, we found longer fixation durations when previews were low frequency words, compared to high frequency words. While the finding is not completely new (see Risse & Kliegl, 2014; Schotter et al., 2012; Schotter, von der Malsburg, & Leinenger, 2019), in previous studies the effects of preview validity and preview difficulty were confounded due to the experimental design. The design of the present experiment allowed for a clear distinction.

Surprisingly, we observed no effect of preview difficulty in fixations on word n—before the boundary. Previous studies (Risse & Kliegl, 2012) had indicated, that lexical properties of parafoveal previews would influence fixation durations already during parafoveal processing, before fixating the target word $n+1$. Such results could indicate interference caused by parallel processing of word n and the preview of word $n+1$. We therefore decided to explore, whether the observed deviant effect pattern could still be explained by frameworks of parallel processing, specifically the SWIFT model, through a simulation study.

As an intermediate step, in Chapter 3, we first implemented a novel technique of parameter estimation for the SWIFT model. For previous simulation studies (e.g., Risse et al., 2014) model parameters were estimated with minimization algorithms, typically based on comparisons of summary statistics of simulated and empirical data. A better approach is to directly calculate the likelihood of individual fixation sequences. Since finding a closed-form solution for the SWIFT model was not feasible, we combined techniques of simulation and approximations to arrive at a stochastic likelihood function. We tested the likelihood function by calculating the likelihood of data simulated by the model, while systematically varying single parameters. In the next step, we estimated the likelihood distribution of four parameters for simulated data within a Bayesian framework, using an adaptive Metropolis-Hastings MCMC algorithm (Vihola, 2012), and successfully recovered the parameters that were used during the simulations. Finally, we estimated four parameters for empirical data from the experiment in Chapter 2 on the participant level. Using the resulting parameter values for simulations produced good agreements between fixation durations and fixation probabilities of experimental data and simulations, respectively.

In Chapter 4 we combined the aspirations of model exploration arising from Chapter 2 with the parameter estimation technique established in Chapter 3. For the new parameter estimations we used only data from the control condition of the experiment in Chapter 2, again on the participant level. Switching to a different algorithm from the same class (previously used in Rabe et al., 2020) allowed us to include more parameters with less data, as well as conducting the estimations in a parallel computing environment, which greatly improved computing time. The resulting posterior parameter distributions were used as sampling pool for data generation.

To explore the possible mechanisms behind the preview difficulty effect, two variants of parafoveal inhibition were implemented into the SWIFT model. In the first variant, processing of the upcoming word in the parafovea immediately influenced the autonomous saccade timer, similar to foveal inhibition. In the second variant, the influence of parafoveal word processing on the saccade timer was delayed by $100ms$. Since introducing such modifications affects the reading process globally, parameters had to be estimated separately for each model variant. We further implemented two variants of saccade cancellation, but as they were strictly coupled to the display change—which was not present in the control condition—they solicited no further estimations. In the first variant, an ongoing saccade program that was still in its labile stage could be cancelled with a probability of 50% at the moment of the display change, during the onset of the first fixation after the boundary. In the second variant, the probability of cancellation was further tied to the first stage of preview processing.

Data were simulated for all three experimental conditions and model variants. We calculated the effect sizes of the preview validity and preview difficulty effects, respectively, and compared them to the experiment results. Models without parafoveal inhibition did not produce an effect of preview difficulty during the first fixation on the target word $n+1$, but yielded small effects of preview validity. When inhibition by parafoveal processing was introduced, difficulty effects increased, but validity effects were not affected. Difficulty effects again slightly increased, when the inhibition by parafoveal words was delayed by $100ms$. The possibility of aborting an ongoing saccade program during its labile stage after display changes was then introduced to imitate a reaction of the oculomotor system to the disturbance of visual continuity by the switch from preview to target word. While this only slightly increased the size of the preview difficulty effect, estimates of the validity effect dramatically increased. When saccade cancellation was restricted to cases, where processing of the preview had not yet advanced to the post-lexical stage, the effects of preview validity were markedly reduced. From these simulations we concluded that the effect pattern observed in Chapter 2 can be explained by the SWIFT model, when parafoveal processing is given a pathway of influencing the autonomous saccade timer, similar to foveal inhibition. While this refinement is sufficient in principle, preview validity effects become most visible when processing dependent saccade canceling is added.

5.2 The order of word processing

The interest in lexical processing of words in the parafovea is strongly related to the debate on the order of word processing, and the connection between cognition and the control of eye movements during reading. Observations of PoF effects in corpus studies (Kennedy & Pynte, 2005; Kliegl et al., 2006) were seen as possible indicators of cross-talk and, hence, parallelism in word processing (Drieghe et al., 2008).

This effect pattern posed a serious problem to SAS models like E-Z Reader, where word processing follows a rigid order and is strictly coupled to saccade programming. In E-Z Reader, finishing the first stage of word processing inevitably triggers a saccade program towards the upcoming word. This architecture provides a mechanism, through which the first processing stage affects fixation durations on the word being currently fixated (i.e., by influencing the onset of saccade programming to the upcoming word), and the second processing stage affects fixation durations on the upcoming word (i.e., by influencing the time of the attention shift to the next word). However, unless saccade cancellation from a competing saccade program is taken into account[1], no mechanism exists that could facilitate an influence of parafoveal processing on the duration of the current fixation.

Experimental investigations using the boundary paradigm eventually lead to the discovery of consistent preview difficulty effects on the word after the boundary (Schotter & Leinenger, 2016; Schotter et al., 2019), or in the upcoming fixation (Risse & Kliegl, 2012), respectively, whereas effects on the word before the boundary were spurious (Hyönä & Bertram, 2004), and the notion of lexical PoF was largely dismissed in a meta-analysis (Brothers et al., 2017).

However, preview difficulty effects on the post-boundary word remained challenging to SAS models. While they allow complete processing of words in the parafovea, the strict coupling of word processing to oculomotor planning would dictate the preparation of a skipping saccade. A solution was found in the forced fixation account. Forced fixations occur, when—during its parafoveal presentation—an upcoming word is processed at least up the point where a new saccade program is triggered, while an ongoing saccade program targeting that word has already transitioned to the non-labile stage of saccade preparation. The ongoing saccade program can, thus, not be canceled in favor of a skipping saccade, resulting in a fixation on the upcoming word. After the onset of the forced fixation the postponed saccade program would go into preparation immediately, or even earlier (Morrison, 1984; Schotter & Leinenger, 2016), resulting in a particularly short fixation. A population of forced fixations in the distribution of fixation durations on the target word effectively decreases aggregate measures like mean fixation durations, if their incidence is high. Thus, forced fixations can account for patterns of reduced fixation durations on the word after the boundary, and predict, that during a trial either the preview, or the target word is processed, but not both.

In contrast, PG models like Glenmore and SWIFT are less restricted by their basic principles, instead it was argued that "parafoveal-on-foveal effects naturally arise as a direct prediction" (Drieghe et al., 2008) from the perspective of parallel processing. The recent rejection of PoF effects (Brothers et al., 2017) raises the question, whether PG models can entertain mechanisms that reproduce the experimentally observed effects of parafoveal preview difficulty on fixation durations after the boundary, while simultaneously precluding influences of parafoveal previews on the current fixation duration. For example, in the Glenmore (Reilly & Radach, 2002, 2006) model all words within the pro-

[1] If processing of the parafoveal word is quick enough to trigger a new saccade program targeting word $n+2$, while the previous saccade program to word $n+1$ has not yet advanced to the non-labile stage, the fixation on word n is prolonged, additionally resulting in skipping of word $n+1$. However, this course of events would be more likely for high frequency previews, and very unlikely for low frequency previews, resulting in a negative effect of preview difficulty.

cessing span simultaneously compete for processing resources, while saccade programming is triggered by thresholds in word processing. This complex pattern elicits PoF effects (Brothers et al., 2017), rendering the combination of mechanisms in the model unfeasible.

Conversely, the decoupling of saccade preparation and word processing in the SWIFT model, together with the constraint of allowing only inhibition by foveated words, prevents parafoveal words from directly affecting fixation durations. Introducing a mechanism that inhibits the autonomous saccade timer by activation of the upcoming, parafoveal word enabled the model to produce parafoveal difficulty effects in a boundary paradigm context. While this also provides a route for evoking PoF effects, adding a delay to parafoveal inhibition shifted the contribution from parafoveal word activation into the upcoming fixation, visible as an increase in the size of the difficulty effect.

Preview validity effects emerged naturally from the implementation of the display change into the model, but they were small compared to the difficulty effects. In order to improve the numerical results of the simulations, an additional mechanism of saccade cancellation was introduced as a reaction of the oculomotor system to the display change. Motivated by saccadic inhibition (Reingold & Stampe, 2004), this resulted in a sizeable increase in fixation durations on the target word during invalid preview conditions. When this saccade cancellation was further coupled to the first word processing stage (i.e., the probability for saccade cancellation was > 0 only when the preview was still in the first processing stage immediately before the display change), the effect patterns resembled the experimental data most closely.

With the addition of both mechanisms, the model was able to reproduce the two distinct main effects of preview validity and preview difficulty on fixation durations on the target word reasonably well, demonstrating that parallel PG models can in principle account for such effect patterns. It should also be noted, that extending model mechanisms to accommodate specific results is a common occurrence in the scope of SAS models as well (e.g., compare the family of E-Z Reader models in Reichle, 2011).

In contrast to our approach, models using serial attention shifts, specifically E-Z Reader, can naturally reproduce the effect patterns via forced fixations, without the necessity of auxiliary modifications, although so far neither qualitative, nor quantitative simulations have been carried out to confirm these predictions. However, in a recent study, Schotter et al. (2019) scrutinized the forced fixations account in an experiment, where they investigated which word was likely read by subjects at the location of the target word. In an $n+1$ boundary paradigm they manipulated plausibility and frequency of previews and target words. They constructed the sentences, such that the plausibility evaluation would be elicited only towards the end of the sentence. For invalid implausible previews of high frequency, they predicted that the subjects would make more regressions, as high frequency previews would elicit more forced fixations. SAS models predict that a word which receives a forced fixation is not processed during that fixation, as the attention already shifted onto upcoming word. Conversely, subjects would make less regressions if they had seen an invalid, but plausible high frequency preview. Indeed, their results support the forced fixation account (see also Schotter, Leinenger, & von der Malsburg, 2018). However, the results do not speak against parallel processing models like SWIFT. In fact, forced fixations happen in SWIFT as well, when a parafoveal word is chosen as a saccade target in light of being currently processed, and processing finishes before the saccadic movement is finalized. Since saccade targets are chosen during the transition from the labile to the non-labile stage of saccade programming, there is considerable time left for parafoveal processing to conclude, especially, when the word is of high frequency. To include forced fixations into simulations in a boundary paradigm, the display change would have to be implemented differently than in the study in Chapter 4. There we actively changed the word frequency to match the target word and reset internal variables for word activation, processing stage and processing completion after the display change to their initial values and, hence, forced the model to process the target word. Conditioning the reset on preview processing would likely elicit forced fixations similar to those proposed for SAS models.

5.3 Bayesian parameter estimation

Measures of reading vary strongly within, and also between subjects. Bayesian parameter estimation presents a viable tool to capture this behavior for process oriented models, enabling model evaluation under realistic conditions.

In the present work the technique of likelihood based Bayesian parameter estimation was applied to a sophisticated model of eye movements during reading, capable of generating complex scanpaths, where fixation durations and fixation locations depend on the previous trajectory. Parameters were estimated using only a subset of data collected in an experiment to estimate a substantial number of parameters on the participant level. Results from cross validation, where the remainder of the participants data was correlated with simulation results based on the parameter estimation, look very promising.

The results can be further improved from using hierarchical Bayesian estimation, or otherwise obtaining more informative priors based on external measures, such as estimates of the processing span or word naming latency (see Kuperman & Van Dyke, 2011), previous estimations or meta-analyses.

While in the current work we utilized the likelihood function, albeit with approximations, this approach is not feasible for some models, specifically when they cannot technically account for all data (i.e., long regression saccades to words $< n - 1$ have zero probability in E-Z Reader, see Reichle et al., 2009). In this case, alternatives based on simulation statistics, like synthetic likelihoods (Wood, 2010) might be a solution.

References

Aarts, E., Verhage, M., Veenvliet, J. V., Dolan, C. V., & Van Der Sluis, S. (2014). A solution to dependency: using multilevel analysis to accommodate nested data. *Nature neuroscience*, *17*(4), 491–496.

Amari, S.-i. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, *27*(2), 77–87.

Amari, S. V., & Misra, R. B. (1997). Closed-form expressions for distribution of sum of exponential random variables. *IEEE Transactions on Reliability*, *46*(4), 519–522.

Andrieu, C., & Roberts, G. O. (2009, 4). The pseudo-marginal approach for efficient monte carlo computations. *The Annals of Statistics*, *37*(2), 697–725.

Angele, B., & Rayner, K. (2011). Parafoveal processing of word n+ 2 during reading: Do the preceding words matter? *Journal of Experimental Psychology: Human Perception and Performance*, *37*(4), 1210–1220.

Angele, B., Slattery, T. J., & Rayner, K. (2016). Two stages of parafoveal processing during reading: Evidence from a display change detection task. *Psychonomic Bulletin & Review*, *23*(4), 1241–1249.

Angele, B., Slattery, T. J., Yang, J., Kliegl, R., & Rayner, K. (2008). Parafoveal processing in reading: Manipulating n+ 1 and n+ 2 previews simultaneously. *Visual Cognition*, *16*(6), 697–707.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Bach, M. (2007). The freiburg visual acuity test-variability unchanged by post-hoc reanalysis. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *245*(7), 965–971.

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, *17*(3), 364–390.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, Articles*, *67*(1), 1–48.

Becker, W., & Jürgens, R. (1979). An analysis of the saccadic system by means of double step stimuli. *Vision Research*, *19*(9), 967–983.

Beer, R. D. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*, *4*(3), 91–99.

Bouma, H. (1973). Visual interference in the parafoveal recognition of initial and final letters of words. *Vision Research*, *13*(4), 767–782.

Brothers, T., Hoversten, L. J., & Traxler, M. J. (2017). Looking back on reading ahead: No evidence for lexical parafoveal-on-foveal effects. *Journal of Memory and Language*, *96*, 9–22.

Brysbaert, M., Drieghe, D., & Vitu, F. (1998). Word skipping: Implications for theories of eye movement control in reading. In *Eye guidance in reading and scene perception* (pp. 125–147). Elsevier.

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.

Coelho, C. A. (1998). The generalized integer gamma distribution—a basis for distributions in multivariate statistics. *Journal of Multivariate Analysis*, *64*(1), 86–102.

Deutsch, A., Frost, R., Pelleg, S., Pollatsek, A., & Rayner, K. (2003). Early morphological effects in reading: Evidence from parafoveal preview benefit in hebrew. *Psychonomic Bulletin & Review*, *10*(2), 415–422.

Ditchburn, R., & Ginsborg, B. (1952). Vision with a stabilized retinal image. *Nature*, *170*(4314), 36–37.

Drieghe, D. (2011). Parafoveal-on-foveal effects on eye movements during reading. In S. E. S. P. Liversedge I. D. Gilchrist (Ed.), *Eye movements handbook* (pp. 839–855). Oxford University Press.

Drieghe, D., Rayner, K., & Pollatsek, A. (2008). Mislocated fixations can account for parafoveal-on-foveal effects in eye movements during reading. *Quarterly Journal of Experimental Psychology*, *61*(8), 1239–1249.

Dutton, J. M., & Starbuck, W. H. (1971). Computer simulation models of human behavior: A history of an intellectual technology. *IEEE Transactions on Systems, Man, and Cybernetics*(2), 128–171.

Engbert, R., & Kliegl, R. (2011). Parallel graded attention models of reading. In S. P. Liversedge, I. Gilchrist, & S. Everling (Eds.), *The oxford handbook of eye movements* (pp. 787–800). Oxford University Press.

Engbert, R., & Krügel, A. (2010). Readers use Bayesian estimation for eye movement control. *Psychological Science*, *21*(3), 366–371.

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.

Engbert, R., & Mergenthaler, K. (2006). Microsaccades are triggered by low retinal image slip. *Proceedings of the National Academy of Sciences*, *103*(18), 7192–7197.

Engbert, R., & Nuthmann, A. (2008). Self-consistent estimation of mislocated fixations during reading. *PLoS One*, *3*(2), e1534: 1–6.

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.

Epanechnikov, V. A. (1969). Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications*, *14*(1), 153–158.

Eriksen, C. W., & James, J. D. S. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, *40*(4), 225–240.

Erlhagen, W., & Schöner, G. (2002). Dynamic field theory of movement preparation. *Psychological Review*, *109*(3), 545–572.

Everatt, J., & Underwood, G. (1994). Individual differences in reading subprocesses: Relationships between reading ability, lexical access, and eye movement control. *Language and Speech*, *37*(3), 283–297.

Findlay, J. M., & Gilchrist, I. D. (2003). *Active Vision: The Psychology of Looking and Seeing* (No. 37). Oxford University Press.

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

Fum, D., Del Missier, F., & Stocco, A. (2007). The cognitive modeling of human behavior: Why a model is (sometimes) better than 10,000 words. *Cognitive Systems Research*, *8*(3), 135–142.

Gardiner, C. (1985). *Handbook of Stochastic Processes.* Springer-Verlag, New York.

Gardner, M. K., Rothkopf, E. Z., Lapan, R., & Lafferty, T. (1987). The word frequency effect in lexical decision: Finding a frequency-based component. *Memory & Cognition*, *15*(1), 24–28.

Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis.* Chapman and Hall/CRC.

Geyken, A. (2007). The DWDS corpus: A reference corpus for the german language of the 20th century. *Collocations and idioms: Linguistic, lexicographic, and computational aspects*, 23–40.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. (1995). *Markov chain Monte Carlo in practice.* Chapman and Hall/CRC.

Gillespie, D. T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, *22*(4), 403–434.

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

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*(1), 97-109.

Heister, J., Würzner, K.-M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., & Kliegl, R. (2011). dlexDB–eine lexikalische Datenbank für die psychologische und linguistische Forschung. *Psychologische Rundschau*, *62*(1), 10–20.

Henderson, J. M., & Ferreira, F. (1990). Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 417–429.

Hohenstein, S., & Kliegl, R. (2014). Semantic preview benefit during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 166–190.

Hohenstein, S., Laubrock, J., & Kliegl, R. (2010). Semantic preview benefit in eye movements during reading: A parafoveal fast-priming study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1150-1170.

Holmes, W. R. (2015). A practical guide to the probability density approximation (pda) with improved implementation and error characterization. *Journal of Mathematical Psychology*, *68-69*, 13–24.

Hyönä, J. (2011). Foveal and parafoveal processing during reading. In S. P. Liversedge, I. Gilchrist, & S. Everling (Eds.), *The oxford handbook of eye movements* (pp. 819–838). Oxford University Press.

Hyönä, J., & Bertram, R. (2004). Do frequency characteristics of nonfixated words influence the processing of fixated words during reading? *European Journal of Cognitive Psychology*, *16*(1-2), 104–127.

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, *40*(6), 431–439.

Inhoff, A. W., Starr, M., & Shindler, K. L. (2000). Is the processing of words during eye fixations in reading strictly serial? *Perception & Psychophysics*, *62*(7), 1474–1484.

Inhoff, A. W., & Tousman, S. (1990). Lexical priming from partial-word previews. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(5), 825–836.

Jordan, T. R., McGowan, V. A., Kurtev, S., & Paterson, K. B. (2016). A further look at postview effects in reading: An eye-movements study of influences from the left of fixation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(2), 296–307.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329-354.

Kennedy, A. (1998). The influence of parafoveal words on foveal inspection time: Evidence for a processing trade-off. In *Eye guidance in reading and scene perception* (pp. 149–179). Elsevier.

Kennedy, A. (2008). Parafoveal-on-foveal effects are not an artifact of mislocated saccades. *Journal of Eye Movement Research*, *2*, 1-10.

Kennedy, A., & Pynte, J. (2005). Parafoveal-on-foveal effects in normal reading. *Vision Research*, *45*(2), 153–168.

Kennedy, A., Pynte, J., & Ducrot, S. (2002). Parafoveal-on-foveal interactions in word recognition. *The Quarterly Journal of Experimental Psychology Section A*, *55*(4), 1307–1337.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*(1-2), 262–284.

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of experimental psychology: General*, *135*(1), 12-35.

Kliegl, R., Risse, S., & Laubrock, J. (2007). Preview benefit and parafoveal-on-foveal effects from word n+2. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(5), 1250–1255.

Krügel, A., & Engbert, R. (2010). The launch-site effect for skipped words during reading. *Vision Research*, *50*, 1532–1539.

Krügel, A., & Engbert, R. (2014). A model of saccadic landing positions in reading under the influence of sensory noise. *Visual Cognition*, *22*(3-4), 334–353.

Kuperman, V., & Van Dyke, J. A. (2011). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of Memory and Language*, *65*(1), 42–73.

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).

Law, K., Stuart, A., & Zygalakis, K. (2015). *Data Assimilation*. Springer.

Lee, H.-W., Legge, G. E., & Ortiz, A. (2003). Is word recognition different in central and peripheral vision? *Vision Research*, *43*(26), 2837–2846.

Legge, G. E., Klitz, T. S., & Tjan, B. S. (1997). Mr. chips: an ideal-observer model of reading. *Psychological Review*, *104*(3), 524–553.

Marin, J.-M., & Robert, C. (2007). *Bayesian Core: A Practical Approach to Computational Bayesian Statistics*. Springer Science & Business Media.

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.

Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin*, *81*(12), 899–917.

McConkie, G. W., Kerr, P. W., Reddix, M. D., & Zola, D. (1988). Eye movement control during reading: I. the location of initial eye fixations on words. *Vision Research*, *28*(10), 1107–1118.

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, *17*(6), 578–586.

Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: evidence for parallel programming of saccades. *Journal of Experimental Psychology: Human*

Perception and Performance, *10*(5), 667–682.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, *47*(1), 90–100.

Niefind, F., & Dimigen, O. (2016). Dissociating parafoveal preview benefit and parafovea-on-fovea effects during reading: A combined eye tracking and eeg study. *Psychophysiology*, *53*(12), 1784–1798.

Nuthmann, A., & Engbert, R. (2009). Mindless reading revisited: An analysis based on the swift model of eye-movement control. *Vision Research*, *49*(3), 322–336.

Nuthmann, A., Engbert, R., & Kliegl, R. (2005). Mislocated fixations during reading and the inverted optimal viewing position effect. *Vision Research*, *45*(17), 2201–2217.

O'Regan, J. K., & Jacobs, A. M. (1992). Optimal viewing position effect in word recognition: A challenge to current theory. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 185-197.

Palestro, J. J., Sederberg, P. B., Osth, A. F., Van Zandt, T., & Turner, B. M. (2018). *Likelihood-free methods for cognitive science*. Springer.

Pollatsek, A., Bolozky, S., Well, A. D., & Rayner, K. (1981). Asymmetries in the perceptual span for Israeli readers. *Brain and Language*, *14*(1), 174–180.

Pollatsek, A., Lesch, M., Morris, R. K., & Rayner, K. (1992). Phonological codes are used in integrating information across saccades in word identification and reading. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 148–162.

Pollatsek, A., Reichle, E. D., & Rayner, K. (2006). Tests of the E-Z Reader model: Exploring the interface between cognition and eye-movement control. *Cognitive Psychology*, *52*(1), 1–56.

R Core Team. (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria.

Rabe, M., Chandra, J., Krügel, A., Seelig, S., Vasishth, S., & Engbert, R. (2020). A Bayesian approach to dynamical modeling of eye-movement control in reading of normal, mirrored, and scrambled texts. *Psychologial Review (in press)*.

Rayner, K. (1975a). Parafoveal identification during a fixation in reading. *Acta Psychologica*, *39*(4), 271–281.

Rayner, K. (1975b). The perceptual span and peripheral cues in reading. *Cognitive Psychology*, *7*(1), 65–81.

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin*, *85*(3), 618–660.

Rayner, K. (1979). Eye guidance in reading: Fixation locations within words. *Perception*, *8*(1), 21–30.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422.

Rayner, K. (2009a). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506.

Rayner, K. (2009b). Eye movements in reading: Models and data. *Journal of Eye Movement Research*, *2*(5), 1–10.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*(3), 191–201.

Rayner, K., Inhoff, A. W., Morrison, R. E., Slowiaczek, M. L., & Bertera, J. H. (1981). Masking of foveal and parafoveal vision during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(1), 167–179.

Rayner, K., & Reichle, E. D. (2010). Models of the reading process. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(6), 787–799.

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1290-1301.

Rayner, K., Well, A. D., & Pollatsek, A. (1980). Asymmetry of the effective visual field in reading. *Perception & Psychophysics*, *27*(6), 537–544.

Rayner, K., Well, A. D., Pollatsek, A., & Bertera, J. H. (1982). The availability of useful information to the right of fixation in reading. *Perception & Psychophysics*, *31*(6), 537–550.

Reich, S., & Cotter, C. (2015). *Probabilistic Forecasting and Bayesian Data Assimilation*. Cambridge University Press.

Reichle, E. D. (2011). Serial-attention models of reading. In S. P. Liversedge, I. Gilchrist, & S. Everling (Eds.), *The oxford handbook of eye movements* (pp. 787–800). Oxford 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.

Reichle, E. D., Rayner, K., & Pollatsek, A. (1999). Eye movement control in reading: Accounting for initial fixation locations and refixations within the ez reader model. *Vision Research*, *39*(26), 4403–4411.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, *26*(4), 445–476.

Reichle, E. D., & Schotter, E. R. (2020). A Computational Analysis of the Constraints on Parallel Word Identification. In *Proceedings of the 42nd annual conference of the cognitive science society.*

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review*, *16*(1), 1–21.

Reilly, R. G., & Radach, R. (2002). Glenmore: An interactive activation model of eye movement control in reading. In *Proceedings of the 9th international conference on neural information processing, 2002. iconip '02.* (Vol. 3, pp. 1194–1200).

Reilly, R. G., & Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. *Cognitive Systems Research*, *7*(1), 34–55.

Reingold, E. M., & Stampe, D. M. (2004). Saccadic inhibition in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(1), 194–211.

Remington, L. A. (2012). Chapter 4 - Retina. In L. A. Remington (Ed.), *Clinical Anatomy and Physiology of the Visual System* (Third Edition ed., p. 61 - 92). Saint Louis: Butterworth-Heinemann.

Risse, S. (2014). Effects of visual span on reading speed and parafoveal processing in eye movements during sentence reading. *Journal of Vision*, *14*(8), 11–11.

Risse, S., Hohenstein, S., Kliegl, R., & Engbert, R. (2014). A theoretical analysis of the perceptual span based on SWIFT simulations of the n+ 2 boundary paradigm. *Visual Cognition*, *22*(3-4), 283–308.

Risse, S., & Kliegl, R. (2012). Evidence for delayed parafoveal-on-foveal effects from word n+ 2 in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 1026–1042.

Risse, S., & Kliegl, R. (2014). Dissociating preview validity and preview difficulty in parafoveal processing of word n+ 1 during reading. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(2), 653–668.

Risse, S., & Seelig, S. (2019). Stable preview difficulty effects in reading with an improved variant of the boundary paradigm. *Quarterly Journal of Experimental Psychology*, *72*(7), 1632–1645.

Robert, C., & Casella, G. (2013). *Monte Carlo Statistical Methods*. Springer Science & Business Media.

Roberts, G. O., Gelman, A., & Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk metropolis algorithms. *The Annals of Applied Probability*, *7*(1), 110–120.

Salthouse, T. A., & Ellis, C. L. (1980). Determinants of eye-fixation duration. *The American Journal of Psychology*, *93*, 207–234.

Schad, D. J., Betancourt, M., & Vasishth, S. (2020). Toward a principled bayesian workflow in cognitive science. *Psychological Methods*.

Schad, D. J., & Engbert, R. (2012). The zoom lens of attention: Simulating shuffled versus normal text reading using the SWIFT model. *Visual Cognition*, *20*(4-5), 391–421.

Schiepers, C. (1980). Response latency and accuracy in visual word recogniton. *Perception & Psychophysics*, *27*(1), 71–81.

Schotter, E. R. (2013). Synonyms provide semantic preview benefit in english. *Journal of Memory and Language*, *69*(4), 619–633.

Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading.

Attention, Perception, & Psychophysics, *74*(1), 5–35.

Schotter, E. R., & Leinenger, M. (2016). Reversed preview benefit effects: Forced fixations emphasize the importance of parafoveal vision for efficient reading. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(12), 2039–2067.

Schotter, E. R., Leinenger, M., & von der Malsburg, T. (2018). When your mind skips what your eyes fixate: How forced fixations lead to comprehension illusions in reading. *Psychonomic Bulletin & Review*, *25*(5), 1884–1890.

Schotter, E. R., von der Malsburg, T., & Leinenger, M. (2019). Forced fixations, trans-saccadic integration, and word recognition: Evidence for a hybrid mechanism of saccade triggering in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(4), 677–688.

Schroyens, W., Vitu, F., Brysbaert, M., & d'Ydewalle, G. (1999). Eye movement control during reading: Foveal load and parafoveal processing. *The Quarterly Journal of Experimental Psychology Section A*, *52*(4), 1021–1046.

Schütt, H. H., Rothkegel, L. O., 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.

Scott, D. W. (2015). *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons.

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.

Shockley, E. M., Vrugt, J. A., & Lopez, C. F. (2017). Pydream: High-dimensional parameter inference for biological models in python. *Bioinformatics*, *34*(4), 695–697.

Sisson, S. A., & Fan, Y. (2011). Likelihood-free Markov chain Monte Carlo. In (pp. 313–335). Chapman & Hall/CRC, New York.

Slattery, T. J., Angele, B., & Rayner, K. (2011). Eye movements and display change detection during reading. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(6), 1924–1938.

Snell, J., & Grainger, J. (2019). Readers are parallel processors. *Trends in Cognitive Sciences*, *23*(7), 537–546.

Snell, J., van Leipsig, S., Grainger, J., & Meeter, M. (2018). Ob1-reader: A model of word recognition and eye movements in text reading. *Psychological Review*, *125*(6), 969–985.

Snell, J., Vitu, F., & Grainger, J. (2017). Integration of parafoveal orthographic information during foveal word reading: beyond the sub-lexical level? *Quarterly Journal of Experimental Psychology*, *70*(10), 1984–1996.

Starr, M. S., & Rayner, K. (2001). Eye movements during reading: Some current controversies. *Trends in Cognitive Sciences*, *5*(4), 156–163.

ter Braak, C. J. (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.

ter Braak, C. J., & Vrugt, J. A. (2008). Differential evolution Markov chain with snooker updater and fewer chains. *Statistics and Computing*, *18*(4), 435–446.

Turner, B. M., & Sederberg, P. B. (2014). A generalized, likelihood-free method for posterior estimation. *Psychonomic Bulletin & Review*, *21*(2), 227–250.

Underwood, N. R., & McConkie, G. W. (1985). Perceptual span for letter distinctions during reading. *Reading Research Quarterly*, *20*(10), 153–162.

Van Essen, D. C. (2003). Organization of visual areas in macaque and human cerebral cortex. *The Visual Neurosciences*, *1*, 507–521.

Van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, *21*(5), 615–628.

Van Kampen, N. G. (1992). *Stochastic Processes in Physics and Chemistry* (Vol. 1). Elsevier: North-Holland.

Vasilev, M. R., & Angele, B. (2017). Parafoveal preview effects from word n+ 1 and word n+ 2 during reading: A critical review and Bayesian meta-analysis. *Psychonomic Bulletin & Review*, *24*(3), 666–689.

Vihola, M. (2012). Robust adaptive metropolis algorithm with coerced acceptance rate. *Statistics and Computing*, *22*(5), 997–1008.

Vitu, F., McConkie, G. W., Kerr, P., & O'Regan, J. K. (2001). Fixation location effects on fixation durations during reading: An inverted optimal viewing position effect. *Vision Research*, *41*(25-26), 3513–3533.

Von der Malsburg, T., & Angele, B. (2017). False positives and other statistical errors in standard analyses of eye movements in reading. *Journal of Memory and Language*, *94*, 119–133.

Wade, N. J., & Tatler, B. W. (2008). Did javal measure eye movements during reading? *Journal of Eye Movement Research*, *2*(5), 1–7.

Wang, Y., Wang, X., & Wu, Y. (2020). A simple model of reading eye movement based on deep learning. *IEEE Access*, *8*, 193757-193767.

White, S. J., Rayner, K., & Liversedge, S. P. (2005). Eye movements and the modulation of parafoveal processing by foveal processing difficulty: A reexamination. *Psychonomic Bulletin & Review*, *12*(5), 891–896.

Wood, S. N. (2010). Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, *466*(7310), 1102–1104.

Yan, M., Richter, E. M., Shu, H., & Kliegl, R. (2009). Readers of chinese extract semantic information from parafoveal words. *Psychonomic Bulletin & Review*, *16*(3), 561–566.

Acknowledgements

I would like to thank my supervisors Sarah Risse and Ralf Engbert for their competent guidance, incredible patience, and warmhearted goodwill. I would not have managed this without their constant encouragement. Their incredible competence and knowledge, the discussions about theories, methods, and results were extremely stimulating.

Further I want to thank all my wonderful colleagues who over the last six years created a great atmosphere for learning and thriving not only in Golm, but also Vienna, Wuppertal, Marburg, Alicante & Loretta. You were a great and inspiring team!

Finally, I also must thank my long time office cohabitant Johan for the calm and company. May we finally play that saxophone-guitar duet some day.