Human scanpaths in natural scene viewing and natural scene search

the role of systematic eye-movement tendencies

Lars Oliver Martin Rothkegel

Day of submission: March 28, 2018

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.

1st reviewer: Prof. Ralf Engbert
2nd reviewer: Prof. Ben Tatler
3rd reviewer: Prof. Felix Wichmann
Day of oral defense: November 12, 2018

Published online at the
Institutional Repository of the University of Potsdam:
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-420005
https://doi.org/10.25932/publishup-42000

Acknowledgements

At this point I would like to thank my two supervisors Hans Trukenbrod and Ralf Engbert who were the best supervisors one could wish for. Whenever I had any doubts, problems or questions, they took their time to help me to improve my work and discuss my problems. I am really glad to have found such competent, nice, and flexible supervisors!

I want to thank Heiko Schütt and Felix Wichmann from Tübingen for working with us throughout the last 3 years. I really enjoyed working in this team and think the cooperation was fruitful in every kind of way, scientifically and personally, and I am looking forward to the next half year of our cooperation!

I want to thank all members of the Eye Lab in Potsdam. Without their help in acquiring the data it would have taken a lot longer to finish this thesis. My special thanks go to Petra Schienmann for coordinating and supervising the Eye Lab all these years!

I want to thank all my colleagues for vivid discussions, nice conference evenings, interesting ideas and last but not least all these lunches in the Mensa. I especially want to thank Anke Cajar for helping me to get all my stuff together for the thesis and for constantly helping me to improve my work. I also want to thank Daniel Backhaus for sharing the office with me and for always cheering me up whenever the mood went slightly below threshold.

Last but not least I want to thank Maria. Just because.

Abstract

Understanding how humans move their eyes is an important part for understanding the functioning of the visual system. Analyzing eye movements from observations of natural scenes on a computer screen is a step to understand human visual behavior in the real world. When analyzing eye-movement data from scene-viewing experiments, the important questions are where (fixation locations), how long (fixation durations) and when (ordering of fixations) participants fixate on an image. By answering these questions, computational models can be developed which predict human scanpaths. Models serve as a tool to understand the underlying cognitive processes while observing an image, especially the allocation of visual attention.

The goal of this thesis is to provide new contributions to characterize and model human scanpaths on natural scenes. The results from this thesis will help to understand and describe certain systematic eye-movement tendencies, which are mostly independent of the image. One eye-movement tendency I focus on throughout this thesis is the tendency to fixate more in the center of an image than on the outer parts, called the central fixation bias. Another tendency, which I will investigate thoroughly, is the characteristic distribution of angles between successive eye movements.

The results serve to evaluate and improve a previously published model of scanpath generation from our laboratory, the SceneWalk model. Overall, six experiments were conducted for this thesis which led to the following five core results:

i) A spatial *inhibition of return* can be found in scene-viewing data. This means that locations which have already been fixated are afterwards avoided for a certain time interval (Chapter 2).

ii) The initial fixation position when observing an image has a long-lasting influence of up to five seconds on further scanpath progression (Chapter 2 & 3).

iii) The often described central fixation bias on images depends strongly on the duration of the initial fixation. Long-lasting initial fixations lead to a weaker central fixation bias than short fixations (Chapter 2 & 3).

iv) Human observers adjust their basic eye-movement parameters, like fixation durations and saccade amplitudes, to the visual properties of a target they look for in visual search (Chapter 4).

v) The angle between two adjacent saccades is an indicator for the selectivity of the upcoming saccade target (Chapter 4).

All results emphasize the importance of systematic behavioral eye-movement tendencies and dynamic aspects of human scanpaths in scene viewing.

Zusammenfassung

Die Art und Weise, wie wir unsere Augen bewegen, ist ein bedeutender Aspekt des visuellen Systems. Die Analyse von Augenbewegungen beim Betrachten natürlicher Szenen auf einem Bildschirm soll helfen, natürliches Blickverhalten zu verstehen. Durch Beantwortung der Fragen wohin (Fixationsposition), wie lange (Fixationsdauern) und wann (Reihenfolge von Fixationen) Versuchspersonen auf einem Bild fixieren, lassen sich computationale Modelle entwickeln, welche Blickspuren auf natürlichen Bildern vorhersagen. Modelle sind ein Werkzeug, um zugrunde liegende kognitive Prozesse, insbesondere die Zuweisung visueller Aufmerksamkeit, während der Betrachtung von Bildern zu verstehen. Das Ziel der hier vorliegenden Arbeit ist es, neue Beiträge zur Modellierung und Charakterisierung menschlicher Blickspuren auf natürlichen Szenen zu liefern. Speziell systematische Blicksteuerungstendenzen, welche größtenteils unabhängig vom betrachteten Bild sind, sollen durch die vorliegenden Studien besser verstanden und beschrieben werden. Eine dieser Tendenzen, welche ich gezielt untersuche, ist die Neigung von Versuchspersonen, die Mitte eines Bildes häufiger als äußere Bildregionen zu fixieren. Außerdem wird die charakteristische Verteilung der Winkel zwischen zwei aufeinanderfolgenden Sakkaden systematisch untersucht.

Die Ergebnisse dienen der Evaluation und Verbesserung des SceneWalk Modells für Blicksteuerung aus unserer Arbeitsgruppe. Insgesamt wurden 6 Experimente durchgeführt, welche zu den folgenden fünf Kernbefunden führten:

i) Ein örtlicher *inhibition of return* kann in Blickbewegungsdaten von Szenenbetrachtungsexperimenten gefunden werden. Das bedeutet, fixierte Positionen werden nach der Fixation für einen bestimmten Zeitraum gemieden (Kapitel 2).

ii) Die Startposition der Betrachtung eines Bildes hat einen langanhaltenden Einfluss von bis zu fünf Sekunden auf die nachfolgende Blickspur (Kapitel 2 & 3).

iii) Die viel beschriebene zentrale Fixationstendenz auf Bildern hängt davon ab, wie lange die erste Fixation dauert. Lange initiale Fixationen führen zu deutlich geringerer zentraler Fixationstendenz als kurze Fixationen (Kapitel 2 & 3).

iv) Menschliche Betrachter passen Fixationsdauern und Sakkadenamplituden an die visuellen Eigenschaften eines Zielreizes in visueller Suche an (Kapitel 4).

v) Der Winkel zwischen zwei Sakkaden ist ein Indikator dafür, wie selektiv das Ziel der zweiten Sakkade ist (Kapitel 4).

Alle Ergebnisse betonen die Wichtigkeit von systematischem Blickbewegungsverhalten und dynamischen Aspekten menschlicher Blickspuren beim Betrachten von natürlichen Szenen.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Vision is the most important and complex sense of humans (Kolb, Whishaw, & Teskey, 2001). Our ability to rapidly discriminate uncountable visual inputs from each other, instantly associate pictures with meaning and recognize a vast amount of familiar faces is simply remarkable. However, all the wonderful things the visual system is capable of, come at a price: the visual system requires by far the largest brain capacity of all our senses (e.g., Purves et al., 1997; Kolb et al., 2001) and yet, the area with high visual acuity is limited to a tiny part of the visual field, called the fovea. The fovea makes up less than 1% of the retinal area (Kandel et al., 2000) and has the highest density of photoreceptors (Polyak, 1941). Thus, the brain capacity needed to decode input from the fovea is a lot larger than the capacity needed for peripheral parts of the visual field with the same size. If the resolution in the periphery would be as high as in the fovea, our brain would be thousands of times larger and weigh up to ten tons (Findlay & Gilchrist, 2003). Because the fovea only covers about 2° of visual angle, we have to constantly move our eyes and head in order to accurately perceive the visual world around us. Vision is thus not simply light being absorbed by the retina to create an image in the brain but is a fascinating interplay of the stimuli which surround us, and ourselves as observers, deciding where to move the eyes next. This interaction of eye movements and the stream of visual data entering the eyes is often referred to as active vision (Findlay & Gilchrist, 2003). Scientists can exploit the fact that our vision is foveated because of the direct link between where our fovea is directed at and the position within the focus of our visual attention (Henderson, 1992). Thus, investigating eye movements is tightly linked to measuring visual attention (Schneider & Deubel, 1995; Deubel & Schneider, 1996; Rayner, Smith, Malcolm, & Henderson, 2009).

Humans display multiple types of eye movements. We make smooth pursuit movements to follow moving objects (Dodge, 1903; Lisberger, Morris, & Tychsen, 1987), vergence movements to follow objects moving in depth (B. Clark, 1936; Mays, 1984), the optical nystagmus to stabilize an image despite of a moving head (Ehlers, 1925) and more

(for an overview see Liversedge, Gilchrist, & Everling, 2011). The two most commonly investigated eye movements are saccades and fixations. Saccades are fast ballistic movements, which are conducted approximately three times a second to shift the fovea from one location to another. During a saccade, perception is strongly suppressed (Matin, 1974; Thiele, Henning, Kubischik, & Hoffmann, 2002). The time between two saccades, when the eyes seem to stand still and visual information is being processed, is called a fixation. During a fixation the eyes never stand completely still. Small fixational eye movements, called drift, tremor and microsaccades can be found within fixations. Without micro-movements the image fades due to retinal adaptation (Riggs, Ratliff, Cornsweet, & Cornsweet, 1953) and attention shifts can be related to fixational eye movements (Engbert & Kliegl, 2003a).

Scientists have been concerned with eye movements in multiple tasks like reading (Rayner, 1998), driving (Mourant & Rockwell, 1972; Land & Lee, 1994), playing sports (Williams & Davids, 1998; Land & McLeod, 2000; Rodrigues, Vickers, & Williams, 2002) and other areas of interest. One of the tasks which has been investigated very thoroughly in the past 30 years is natural scene viewing (Henderson & Hollingworth, 1998, 2003; Rayner et al., 2009).

The next sections will introduce the scene-viewing paradigm and provide an overview of what is known so far about eye-movement behavior in natural scene viewing. Afterwards, I will describe how computational models predict fixation locations, durations and complete eye traces (so-called scanpaths) of an observation. The SceneWalk model of scanpath generation, which simulates human eye traces on pictures, will be described in detail (Engbert, Trukenbrod, Barthelmé, & Wichmann, 2015). This thesis contains three studies which were designed to evaluate and improve the SceneWalk model. The primary aim is to increase our understanding of the underlying processes for human eye-movement behavior and integrate them into the SceneWalk model. Thus, all studies contain new findings to describe human eye traces from scene-viewing experiments and introduce new ideas for why we move our eyes in systematic ways when looking at a picture.

1.1 The scene-viewing paradigm

Scene-viewing experiments are an intermediate stage between highly controlled laboratory experiments (for example psychophysical contrast discrimination experiments) and the real-world. We are interested in understanding real-world behavior and thus investigate eye movements from scene-viewing experiments as a step toward understanding real-world eye-movement behavior.

In laboratory-based scene-viewing experiments, participants sit in front of a computer screen while their eyes' locations are being recorded with an eyetracker. Before each

experimental trial participants fixate a marker on the screen. This fixation check assures that the eyetracker is recording and participants fixate where they are told to. After the marker is fixated, an image appears. These images are usually photographs which represent an excerpt of the real world.

The task, which participants are required to perform while watching the image, plays an important role, because it has shown to influence the way observers look at the image (Yarbus, Haigh, & Riggs, 1967; Mills, Hollingworth, Van der Stigchel, Hoffman, & Dodd, 2011). The most commonly used tasks are visual search, free viewing and memorization. Other tasks include aesthetic valuation of the image (Nuthmann & Henderson, 2010) or estimating the age of displayed persons (Yarbus et al., 1967).

Besides the task, another important aspect of scene-viewing experiments is the stimulus material. Photos require a very high definition, such that the retinal image of the stimulus is as equal as possible to the real-world excerpt of the scene. To acquire as much equality as possible between the real-world excerpt and the photograph, sharpness should be uniformly high within the picture. Unfortunately, an image with equal sharpness at every position can not represent the depth of sharpness the eyes create in a natural environment, but at least the input at the fixation position is continuously sharp.

Besides the images, properties of the monitor and the laboratory setting play an important role to recreate the visual impression of the real-world environment where the image was taken (e.g., the definition of the monitor must be high enough and brightness must be uniform).

Even if an experiment is designed with all required aspects in mind, real-world natural viewing is not equivalent to looking at static photos on a computer screen. How results from lab-based scene viewing transfer to viewing behavior in the natural environment is beyond the scope of this thesis and subject to another field of research (see Hayhoe & Ballard, 2005; Tatler, Hayhoe, Land, & Ballard, 2011). New tools like mobile eye tracking glasses and virtual reality devices will help to bridge the gap from the laboratory to the real world (e.g., 't Hart et al., 2009; Rosa et al., 2015; Binaee, Diaz, Pelz, & Phillips, 2016; Engbert, Rothkegel, Backhaus, & Trukenbrod, 2016).

Once a scene-viewing experiment is designed and eye-tracking data are collected, three main aspects of the eye movements are of general interest. The first aspect is where people fixate. The second is how long they fixate. The third is when (i.e., in what order) they fixate certain locations. A single eye trace on an image, which combines fixation locations and their order, is called a scanpath.

To predict scanpaths in scene-viewing experiments, it is crucial to know where participants fixate. However, it has been shown that knowing where participants look at in an image is not enough to produce valid scanpaths, because each fixation depends on the properties of previous fixations (Foulsham & Underwood, 2008; Tatler & Vincent,

2008; Schütt et al., 2017). Dependencies between successive fixations and saccades can result from memory, oculomotor and spatial attentional preferences, or simply the degradation of visual acuity into the retinal periphery. Before presenting the studies conducted for this thesis, the upcoming sections will provide an overview of what is known about eye-movement behavior in scene viewing.

1.2 Fixation locations in scene viewing

Early work on eye movements in scene viewing showed that human fixation locations are not placed randomly in a scene, but tend to cluster at certain image locations (Buswell, 1935; Yarbus et al., 1967). These findings led to the idea that fixation locations on an image can be predicted. The factors, which contribute to similarities in fixation location placement between different observations, have been divided into three categories. First, bottom-up influences which depend merely on the visual stimulus. Second, top-down influences which depend on the mental state of the observer while looking at the image. Third, systematic eye-movement tendencies, which are mostly independent of image content and can be found in almost all scene-viewing experiments.

1.2.1 Bottom-up influences

A thoroughly studied aspect of where people look in natural scenes are bottom-up influences (e.g., Itti, Koch, & Niebur, 1998; Itti & Koch, 2000; Peters, Iyer, Itti, & Koch, 2005; Judd, Ehinger, Durand, & Torralba, 2009), which emanate from the visual stimulus and thus are driven externally. A part of the bottom-up influences are believed to reflect an evolutionary preference of the human visual system for certain statistical aspects of an image (Itti & Koch, 2001). These so called low-level features[1] are believed to guide the eyes through a scene, by attracting visual attention regardless of higher-level processes (e.g., Treisman & Gelade, 1980; Itti & Koch, 2000). Postulated low-level features are luminance contrast, color contrast and orientation, since cells within the primary visual cortex fire in reaction to specific orientations and spatial frequencies (Blakemore & Campbell, 1969; De Valois, Albrecht, & Thorell, 1982; Schütt & Wichmann, 2017) and the very early visual system (e.g., retinal ganglion cells) is tuned to react to luminance and color differences (e.g., Polyak, 1941; Purves et al., 1997; Kolb et al., 2001). Since low-level features are supposed to involuntarily guide gaze due to properties of the early visual system, they should be found within all healthy observers. Some stimuli, like flashing lights have been shown to attract eye movements, even when participants where asked

[1]The term level refers to the stage within the brain. Low-level stages of the visual system reach from the sensory receptors to the primary visual cortex, high-level refers to later cortical stages like the inferior temporal cortex or the occipital face area.

not to look at them, which is an indication that low-level features may guide gaze irrespective of top-down attentional processes (Hallett, 1978; Theeuwes, Kramer, Hahn, & Irwin, 1998). By correlating fixation positions with image statistics, it has additionally been shown that fixated regions have higher contrast than non-fixated regions (Reinagel & Zador, 1999; Tatler, Baddeley, & Gilchrist, 2005). This is particularly evident when the incoming saccade has a small amplitude (Tatler, Baddeley, & Vincent, 2006).

Besides low-level features, mid-level features, like edge-content with high-spatial frequency (Tatler et al., 2005) or local symmetry (Privitera & Stark, 2000; Kootstra, de Boer, & Schomaker, 2011) and higher-level features like objects (Einhäuser, Spain, & Perona, 2008; Nuthmann & Henderson, 2010) also belong to bottom-up image-based influences which correlate positively with fixation positions. It is rather difficult to completely separate high-level bottom-up influences from top-down influences on fixation selection. Throughout this thesis I will categorize all image-based influences on fixation selection as bottom-up influences (also see Einhäuser et al., 2008; Schütt, Rothkegel, Trukenbrod, Engbert, & Wichmann, 2018). It has been suggested that higher-level objects are more important for fixation selection than low-level features (Einhäuser et al., 2008; Nuthmann & Henderson, 2010). Although results of the study by Einhäuser et al. (2008) have been questioned (Borji, Sihite, & Itti, 2013), the notion that object features predict fixations better than low-level features was confirmed in the recent past by very sophisticated modeling techniques (Kümmerer, Wallis, & Bethge, 2016; Schütt et al., 2018). The bottom-up approach for defining where humans look at has been rather successful, but even all low-, mid-, and high-level bottom-up features combined cannot explain all variance in fixation locations because top-down factors are also rather important for understanding where humans look at in pictures (e.g., Oliva, Torralba, Castelhano, & Henderson, 2003; Navalpakkam & Itti, 2005).

1.2.2 Top-down influences

Top-down influences are observer and task dependent and become apparent when eye-movement data from observations on the same image but under different conditions are investigated (e.g., Yarbus et al., 1967; Nuthmann & Henderson, 2010; Mills et al., 2011). A widely known study by Yarbus (1967) has shown that the instruction given to a participant (e.g., estimate the age of the people in the picture vs. remember the clothes of the people) strongly influences where they look at. Regardless of the fact that only one person performed this task under rather unnatural conditions (see DeAngelus & Pelz, 2009), the influence of the task on fixation locations has been replicated multiple times

since (e.g., Neider & Zelinsky, 2006; Castelhano, Mack, & Henderson, 2009; Borji & Itti, 2014). Additionally, it has been shown that when non-salient image regions become important for the viewing task at hand, participants fixate these regions or even produce a characteristic scanpath pattern on a blank screen, suggesting that top-down influences can completely take over (Ferreira, Apel, & Henderson, 2008; Einhäuser, Rutishauser, Koch, et al., 2008; Henderson, 2011). In natural-environment tasks (e.g., tea and sandwich making) with mobile eye tracking, it has also been shown that the eyes rather focus on task-related than salient objects (Land & Hayhoe, 2001; Hayhoe & Ballard, 2005). Top-down factors, other than the task, include the observers memory and scene understanding. If two objects in a scene contain equal visual saliency, but one object is less predictable (e.g., an octopus on a farm), this surprising object receives a higher amount of fixations (Loftus & Mackworth, 1978) than the predictable object (e.g., a tractor on a farm).

Top-down factors are especially important in visual search tasks. Many studies of visual search have shown that a target template is stored in the mind to guide gaze through a scene. This template has shown to guide the eyes in a feature-based (e.g., looking at all red objects when searching a red rose; see Wolfe, 1994; Hwang, Higgins, & Pomplun, 2009) and context-based manner (e.g., looking at the sky when searching for a helicopter; see Henderson, Weeks Jr, & Hollingworth, 1999; Oliva et al., 2003; Torralba, Oliva, Castelhano, & Henderson, 2006; Neider & Zelinsky, 2006). Another factor which guides the eyes in visual search is the expected value associated with a target, thus eye-movements are conducted in a way to increase possible reward (Navalpakkam, Koch, Rangel, & Perona, 2010; Tatler et al., 2011).

Thus, top-down factors like the task, scene understanding and expected reward modulate viewing and search behavior on scenes and cannot be disregarded when investigating human eye traces.

1.2.3 Systematic tendencies

Besides bottom-up and top-down influences, it has been shown that statistical regularities in scanpaths, so called systematic eye-movement tendencies, can explain a large amount of variance in human eye-movement behavior (Tatler & Vincent, 2008, 2009). These systematic tendencies are very persistent and difficult to be experimentally eliminated from human eye movement behavior during scene viewing. Analyzing systematic eye-movement tendencies plays a major role in determining fundamental rules of how humans move their eyes. Examples for spatial behavioral biases are the tendency to make more horizontal than vertical and more vertical than oblique saccades (e.g., Bair & O'keefe, 1998; Tatler & Vincent, 2009; Bays & Husain, 2012) - at least on images which are not tilted (Foulsham, Kingstone, & Underwood, 2008) -, the tendency to make short

saccades (Bahill, 1975; Tatler et al., 2006), and the tendency to fixate more on the left than on the right side of an image (Abed, 1991; Ossandón, Onat, & König, 2014). Two well-known tendencies will be treated thoroughly throughout this thesis. The first one is the central fixation bias, which behaviorally manifests in fixating more often in the center of an image than close to the borders (e.g., Buswell, 1935; Tatler, 2007). This bias is so strong, that it is the single best predictor for fixation locations in scene viewing (Vincent, Baddeley, Correani, Troscianko, & Leonards, 2009; Judd et al., 2009). The second one is the distribution of angles between two successive saccades. Most saccades follow the same direction as the previous saccade or completely reverse direction (Smith & Henderson, 2009; Luke, Schmidt, & Henderson, 2013; Wilming, Harst, Schmidt, & König, 2013). Because the distribution of angles between successive saccades has been shown to be consistent over many scene-viewing experiments, we tried to find a meaningful explanation for this behavior.

A main purpose of this thesis was to find out more about when systematic eye-movement tendencies occur and why they exist in natural scene viewing. Because they are so common and reliable, implementing systematic tendencies into models of eye movement control improves model performance substantially (Tatler & Vincent, 2009; Le Meur & Liu, 2015).

1.3 Fixation durations in scene viewing

Fixation durations are an important measure for predicting and understanding human eye-movement behavior because they reflect ongoing cognitive processes (Rayner et al., 2009). Fixation duration distributions from scene-viewing experiments are typically right-skewed with a mean of about 300 ms (e.g., Henderson & Hollingworth, 1998; Henderson, 2011) and large deviations both within and between subjects (Henderson & Hollingworth, 2003; Castelhano & Henderson, 2008b).

Many scientists have tried to answer which factors determine the duration of fixations. To keep categorization throughout this thesis congruently, I will further group influences on fixation durations, as for fixation locations, into bottom-up, top-down and systematic tendencies. However, in the past it has been discussed if fixation durations underly *direct control, indirect control* or *mixed control* (for an overview see Henderson & Smith, 2009; Trukenbrod & Engbert, 2014). Unfortunately, direct and indirect control of fixation duration are not completely congruent to bottom-up and top-down control. Direct control means that the current visual input during a fixation is responsible for the length of this fixation (e.g., Rayner, 1995). Indirect control means that a fixation duration does not depend on the current foveal input. An example of indirect control would be an autonomous timer, which ends the fixation by triggering a saccade after a random time

interval (Engbert & Kliegl, 2001; Nuthmann, Smith, Engbert, & Henderson, 2010). In mixed control, a combination of indirect and direct factors determine fixation durations (e.g., Henderson & Pierce, 2008).

1.3.1 Bottom-up influences

To investigate how fixation durations depend on the visual input, experimenters use different types of image manipulations, like changing the whole image with respect to low-level feature content. If fixations differ between manipulated images and the original images, this manipulation has a so-called global effect on fixation durations, since it changes the overall distribution of fixation durations. Examples for whole-image manipulations, which have led to increased fixation durations in scene viewing compared to control experiments, are luminance reduction (Loftus, 1985), spatial frequency filtering (S. Mannan, Ruddock, & Wooding, 1995) and color removal (Ho-Phuoc, Guyader, Landragin, & Guérin-Dugué, 2012).

Experiments, where manipulations of the scene happen only in a certain part of the visual field have also produced global changes in fixation duration distributions. For example, spatial frequency filtering led to increased fixation durations compared to unfiltered images, when low-spatial frequencies were removed from the central visual field and high-spatial frequencies were removed from the periphery (Laubrock, Cajar, & Engbert, 2013).

An experimental design to investigate bottom-up influences on individual fixation durations is the Scene-Onset-Delay paradigm (SOD Shioiri, 1993). In SOD experiments, a mask is placed over the scene for a variable time interval during a crucial saccade before the original scene is restored. Results from SOD experiments have shown that manipulations can directly influence the duration of the critical fixation (i.e., the fixation which coincides with the mask). Two different distributions of durations were found for the critical fixations. One was prolonged by the length of the delay interval and one was unaffected by the manipulation (Henderson & Pierce, 2008). This was seen as an indiction that the current visual input and indirect factors determine fixation durations.

Recently, Nuthmann (2017) used linear mixed models to investigate the influence of scene statistics at the position of gaze on fixation durations. Results have shown that luminance and contrast correlated negatively with fixation durations, i.e., areas with low contrast and luminance had higher fixation durations. The amount of edge and object information at the current fixation location correlated positively with fixation durations. Another result from this study was that visual features of a current fixation position

interact with the preceding and successive fixation duration. As for fixation locations, bottom-up influences like low- mid- and high-level image features influence fixation durations but cannot explain all variance.

1.3.2 Top-down influences

The fact that fixation durations are a measure of ongoing cognitive processes suggests that they depend on the observers state of mind. As for fixation locations, one of the cognitive influences on fixation duration, which has been investigated thoroughly, is the viewing task. However, the question whether viewing task influences average fixation durations has not been answered consistently. Castelhano and colleagues (Castelhano et al., 2009) did not find a significant difference between a memorization and a search task in a scene-viewing experiment whereas other studies found a task dependency on fixation durations, such that in visual search fixation durations are shorter than in memorization (Henderson & Hollingworth, 1999; Võ & Henderson, 2009; Mills et al., 2011; Nuthmann, 2017).

Another top-down factor, which I have also mentioned for fixation locations, is the scene context. For example, if an object is placed in a rather untypical location or scene (De Graef, Christiaens, & d'Ydewalle, 1990; Loftus & Mackworth, 1978), fixation durations on this object increase compared to typical objects, even when observers did not realize that they have looked at this untypical object (T. H. Cornelissen & Võ, 2017). Top-down influences, like the viewing task and the scene context thus are an important factor for shaping fixation duration distributions.

1.3.3 Systematic tendencies

Systematic tendencies not only influence fixation locations but also fixation durations. For example, fixation durations are strongly influenced by the change in saccade direction, that is, the angle between the incoming and outgoing saccade from a fixation (Nuthmann, 2017; Tatler, Brockmole, & Carpenter, 2017). The influence of the angle on fixation duration has partially been attributed to inhibition of return (Smith & Henderson, 2009), because saccades back to the previous fixation location are preceded by rather long fixations. However, two results promote the idea that the influence of saccadic angle could originate from the oculomotor system instead of a mere attentional inhibition of return mechanism. First, longer fixations appear when saccades change direction but the saccade target does not land on the previous fixation location (Luke, Nuthmann, & Henderson, 2013). Second, fixation duration increases monotonically with increasing angle between the incoming and outgoing saccade (Tatler & Vincent, 2008; Luke, Smith, Schmidt, & Henderson, 2014).

Another systematic influence on fixation durations is viewing time. Fixation durations increase throughout a trial, whereas saccade amplitudes decrease throughout a trial (Antes, 1974; Over, Hooge, Vlaskamp, & Erkelens, 2007; Mills et al., 2011). This behavior is interpreted as the coarse-to-fine eye movement strategy in scene viewing and visual search. This strategy means that an image is initially scanned and observed rather globally (i.e., large saccades, short fixations) before observers try to extract finer details (i.e., small saccades, long fixations).

The fact that influences on fixation locations and fixation durations have much in common, suggests that they are not independent from each other. In agreement with this notion, it has been shown that positions which are fixated often also receive long fixation durations (Einhäuser & Nuthmann, 2016).

1.4 Computational models of eye movements in scene viewing

The past sections have shown that eye-movement data from scene-viewing experiments contain many regularities. These regularities can be used to predict human eye-movement behavior by formulating computational models. The next section will first provide an overview on fixation location and fixation duration models and then introduce a series of dynamical models for scanpath generation.

1.4.1 Models of fixation locations in scene viewing

Within the past 20 years the question of where humans look at in pictures has been subject to a vast amount of research which led to numerous models predicting human fixation locations on images. The first numerical and most influential model to predict fixation locations on any given image was developed by Itti, Koch, and Niebur (1998). This model was based on the feature integration theory (Treisman & Gelade, 1980) which states that a series of single low-level features (e.g., shape & color) can be processed in parallel, irrespective of top-down processes. The model by Itti et al. uses the idea by Koch and Ullman that the preattentively processed features are combined to one single map, which is responsible for attentive selection of fixation locations (Koch & Ullman, 1985). Thus, the model computes allocation of conspicuity within an image in a purely low-level feature-based manner. In the model, low-level feature maps from any given image are computed and combined to create a single 2-D *saliency map*. The features which create the saliency map are local differences in luminance, color and orientation, which correspond to the sensitivity of neurons in the retina, the lateral geniculate nucleus and the primary visual cortex (e.g., Leventhal, 1991; Itti et al., 1998; Schütt & Wichmann,

2017). This follows the logic that positions which are conspicuous on the basis of low-level image features attract human gaze (see Section 1.2.1).

After the original Itti et al. model, which had strong theoretical implications, but performed only slightly above chance in predicting empirical fixation locations, many fixation location models followed, which also produce saliency maps (e.g., Kienzle, Wichmann, Franz, & Schölkopf, 2006; Harel, Koch, & Perona, 2007; Bruce & Tsotsos, 2009). The development of saliency models has become a real competition and new models are developed continuously (for a review, see Borji & Itti, 2013). A website ranking these models has even been established (Bylinskii et al., 2015). Although the original term visual saliency described a combination of local differences in low-level features, which resemble the receptive field properties of the early visual system, by now the majority of stimulus-based fixation location models (or bottom-up models) are referred to as saliency models.

Some newer models, which go beyond the original low-level bottom-up saliency model, take higher-level features like faces or top-down factors like scene context or task into account (Navalpakkam & Itti, 2005; Torralba et al., 2006; Cerf, Harel, Einhäuser, & Koch, 2008). The currently best performing saliency model is the DeepGaze II model (Kümmerer et al., 2016). DeepGaze II uses a deep neural network architecture to extract a large amount of visual features and weights to predict fixation locations. The features from DeepGaze II specifically predict where objects are placed in the scene and which image features are found in objects. The fact that predicting fixation locations by predicting where objects are placed works rather well agrees with previous studies, showing that objects predict eye movements better than low-level saliency (Einhäuser et al., 2008; Stoll, Thrun, Nuthmann, & Einhäuser, 2015; Schütt et al., 2018).

Although newer saliency models perform quite well in predicting the overall placement of fixation locations in free-viewing experiments, they are mechanistically implausible because they are static and do not take the inhomogeneous distribution of photoreceptors in the retina into account. Static models treat an image as an homogeneously perceived stimulus, although different fixation locations create different percepts for the observer. Furthermore, only recently evaluation methods have been established to adequately compare saliency models with a principled metric (Kümmerer, Wallis, & Bethge, 2015; Schütt et al., 2017; Nuthmann, Einhäuser, & Schütz, 2017). Studies providing new evaluation methods for models show that most models rely heavily on the implementation of a central fixation bias (Clarke & Tatler, 2014; Kümmerer et al., 2015; Nuthmann, 2017; Schütt et al., 2018). This means that most models have to heighten activities at the center of the image compared to parts close to the image borders. It has actually been shown no other feature predicts fixations locations as well as the central fixation bias (Judd et al., 2009; Vincent et al., 2009). The central fixation bias exists independently of top-down

or bottom-up processes and thus hides attentional processes elicited by the image or the task. Due to the large influence of the central fixation bias on fixation selection, Chapter 3 of the present thesis will explicitly investigate it's underlying cause.

Most saliency models were designed to predict fixation locations in free-viewing tasks and perform rather poorly on data from visual search experiments (Henderson, Brockmole, Castelhano, & Mack, 2007; Schütt et al., 2018). One way to improve performance of static models in visual search tasks is to take properties of the search target into account, to represent a mentally stored target template (Wolfe, 1994; Navalpakkam & Itti, 2005; Hwang et al., 2009).

1.4.2 Models of fixation durations in scene viewing

In reading, numerical models which predict fixation durations have been established for two decades (Reichle, Pollatsek, Fisher, & Rayner, 1998; Engbert, Nuthmann, Richter, & Kliegl, 2005). In scene viewing these models emerged later, probably because the common interest rather focused on where people look instead of how long they dwell on each location.

The first computational model of fixation durations in scene viewing was the CRISP model (Nuthmann et al., 2010). The model simulates fixation durations by implementing a random-walk. Once a certain threshold is reached by the random walk the model starts a saccade program. The idea of a random walk for fixation duration modeling has also proved its worth in reading models (Engbert & Kliegl, 2003b). In the CRISP model, the random walk can be inhibited by ongoing cognitive processes, i.e., influenced by the current visual input. The random walk increases with a fixed rate, which can be reduced through inhibition by the current visual input. Due to variability between realizations of the random walk, the model can account for the variance in fixation durations. The mixture of an autonomous timer and ongoing cognitive processes is an example of mixed control for fixation durations. Another principle of the CRISP model is that it has two stages of saccade programming. The first one is the labile stage, where a saccade to a certain location is planned but can still be aborted (Becker & Jürgens, 1979; Reichle et al., 1998; Engbert, Longtin, & Kliegl, 2002). The second is the non-labile stage, in which a saccade program can not be canceled anymore.

One shortcoming of the CRISP model, although it produced rather accurate distributions of fixation durations, was that it does not differentiate between peripheral and foveal input, i.e., does not take the inhomogeneity of the retina into account. Thus, Laubrock, Cajar and Engbert (2013) extended the CRISP model by postulating two random-walks, one for peripheral and one for foveal input. With this model they were able to account for a wider range of experimental manipulations such as spatial frequency filtering in selected

areas of the visual field.

Another model, which uses an autonomous timer and inhibition by the current visual input, is the ICAT model by Trukenbrod and Engbert (Trukenbrod & Engbert, 2014). Besides local control, i.e. the current input, fixation durations in this model depend on global control, like the task of the observation and overall processing difficulties within a task, independent of each current fixation properties.

Recently, a new fixation duration model was published by Tatler, Brockmole, and Carpenter (2017). Since this model not only predicts fixation durations but also fixation locations and dependencies between successive fixations, it will be discussed in the next section on dynamical models in scene viewing.

1.4.3 Dynamical models of scanpath generation in scene viewing

To overcome the mechanistic implausibility of static saliency models and incorporate the statistical dependencies between successive eye-movements, dynamical models have been developed to predict complete scanpaths on an image. As afore mentioned, a scanpath is characterized by an ordered series of fixations within an image (Noton & Stark, 1971; Foulsham & Underwood, 2008). The sequential order of fixations and dependencies between successive eye-movements can provide information about the influences of low-level and high-level bottom-up factors on fixation selection (Parkhurst, Law, & Niebur, 2002; Schütt et al., 2018), reveal properties of covert attention (Bays & Husain, 2012; Wilming et al., 2013; Cajar, Schneeweiß, Engbert, & Laubrock, 2016) or even help to predict participants intelligence (Hayes & Henderson, 2017). Leaving out the dynamic aspects of eye movements in scene viewing means neglecting a major aspect of human viewing behavior. Furthermore, it has been shown that by only weighting a saliency map by the distance to a current fixation location, model performance can be improved substantially (Parkhurst et al., 2002). Additionally, modeling eye movements by only using dependencies between successive fixations and saccades outperformed classical saliency modeling (Tatler & Vincent, 2009). Not many dynamic models have been published yet, but the ones available outperform static models substantially (Le Meur & Liu, 2015; Schütt et al., 2017).

Itti and Koch (2000) provided a mechanism to guide gaze through their saliency map. In their model, dynamics are implemented rather simplistically with an inhibition of return mechanism that inhibits recently fixated locations (Posner, Rafal, Choate, & Vaughan, 1985; Klein, 2000). Fixation locations are chosen in a winner-takes-all manner, such that the highest saliency value receives the first fixation. The following fixations are ordered by descending saliency value. This is rather implausible, because no spatial dependencies of successive eye movements are captured by this model. The eye movement system produces saccadic errors, which is another reason why a deterministic fixation selection is

biologically implausible (Deubel, Wolf, & Hauske, 1984; Krügel & Engbert, 2014). Other contradictory results to the postulation of a deterministic scanpath on an image is that participants fixation locations on the same image are highly variable (Castelhano & Henderson, 2008b; Henderson & Luke, 2014) and even scanpaths of the same participant on the same image when viewed a second time are different from first observations (Kaspar & König, 2011; Trukenbrod, Barthelmé, Wichmann, & Engbert, 2017). Thus, the Itti and Koch model, although able to predict overall fixation locations above chance, creates scanpaths which are rather different than those from human observers (Foulsham & Underwood, 2008). Nonetheless, the idea that inhibition of return drives the eyes through a spatial map has been used many times since and Itti and Koch were the first ones to implement it in an eye-movement model for scene viewing.

Newer dynamic models of saccade generation have attempted to implement systematic tendencies and dependencies between successive eye movements. Le Meur (2015) proposed a model where static saliency (he used the GBVS model by Harel et al. (2006) as the static saliency map) is multiplied with the conjunctive probability distribution of saccade amplitude and direction to create a target selection map. Thus, each fixation position leads to a unique target selection map. The distribution of saccade amplitudes and direction is computed from experimental eye-movement data. The model uses a transient inhibition of return for guiding the eye through the image and instead of a winner-takes-all mechanism, the successive fixation is drawn randomly from the 5 positions with the highest activations on the target map. The inhibition of return for each fixation remains active for the following five fixations with decreasing influence for each new fixation. This rather simple adjustment of the dynamic control and the avoidance of a deterministic fixation selection makes the model by Le Meur mechanistically more plausible than the Itti and Koch model, reproduces saccade amplitude and angle distributions and scores higher in standard metrics for measuring performance of saliency models.

One recently published model, the LATEST model by Tatler, Brockmole and Carpenter (2017), predicts scanpaths and fixation durations. The idea behind this model is that humans continuously decide if they want to move the eyes (go) or maintain fixation (stay; see Findlay & Walker, 1999). Previously it has been shown that locations, which are fixated more often, are also fixated longer (Einhäuser & Nuthmann, 2016). This is congruent with the idea that fixation duration and target selection probability are not independent from each other. In the LATEST model, not only the current fixation position influences the stay or go decision, but all possible saccade targets on the image influence how long a fixation duration lasts. Each single location on the map influences a so called LATER unit (based on the LATER model by Noorani & Carpenter, 2016), which accumulates evidence for the decision to start a saccade. As soon as any location on the image has reached a certain threshold, a saccade toward this position is triggered.

Thus, in the model the question of where participants move their eyes depends completely on when they move them. How each location in the image influences fixation duration is predicted by multiple image features, the ordinal fixation number and oculomotor factors like the change in saccade direction. The image features are devided between influences from the current and all possible upcoming fixation locations. Additional to the main part of the distribution of fixation durations, the LATEST model has a so called maverick unit of rather short fixations, which might be either be fixations which are interrupted by pre-programmed saccades or slightly erroneous fixations, which are followed by corrective saccades. Results produced by the LATEST model are promising but some aspects, like the peak of saccades in the opposite direction than the previous one, are not captured by the model, since saccades in the opposite direction than the previous precede rather long fixations but appear very often.

1.4.4 The SceneWalk model of scanpath generation

The SceneWalk model is a dynamical model of scanpath generation (Engbert et al., 2015). I will describe the model in detail, because all three studies within this thesis had the particular goal to improve and evaluate the SceneWalk model. In the studies of this thesis, we used the model to simulate data and added extensions to the model, to reproduce systematic eye-movement tendencies.

The SceneWalk model is based on two different neural activation maps, an attention map which represents allocation of bottom-up visual attention and the inhomogeneity of visual acuity, and a fixation map, which keeps track of visited locations. These two maps are combined to create a target selection map. This assumption of an allocentric dynamic map is supported by physiological studies (Killian, Jutras, & Buffalo, 2012).

The attention map A_{ij} is driven by early visual processing and controls the distribution of visual attention. The attention map is computed by weighting the empirical density map of fixations (computed from experimental data, see Barthelmé, Trukenbrod, Engbert, & Wichmann, 2013) with a two-dimensional Gaussian around the current fixation position (i, j). This Gaussian for a gridded map at each of $k \times l$ positions is given by

$$G_A = \frac{1}{2\pi\sigma_a^2} e^{-\frac{(k-i)^2+(l-j)^1}{2\sigma_a^2}}, \tag{1.1}$$

where σ_a is the standard deviation of the Gaussian. The attention map A_{ij} thus heightens activations close to the current fixation position and reduces activations on positions which are far away from the current fixation position. The motivation behind this is that visual acuity decreases away from the fixation position. Also, visual attention is linked to the foveal position (Henderson, 1992) and is thus generally reduced with increasing distance to the fixation position. The attention map is initialized as the empirical saliency map

multiplied with the two dimensional Gaussian around the starting position.

The second map, the fixation map, keeps track of fixated positions. The fixation map also weights activation around the current fixation position with a two dimensional Gaussian (equivalent to Eq. 1.1 but with a different standard deviation σ_f). The fixation map starts out as zero activation in the original model (Engbert et al., 2015) or constant activation in a newer version of the model (Schütt et al., 2017). At each time-point t within a trial, the attention map and fixation map are updated. The updating formula for the attention map $A(t)$ is

$$A(t) = \frac{G_A \times \phi}{\sum G_A \times \phi} + e^{-\omega_a(t-t_0)} \left(A(t_0) - \frac{G_A \times \phi}{\sum G_A \times \phi} \right), \tag{1.2}$$

where G_A represents the Gaussian map from Equation 1.1 around the current fixation position. $A(t_0)$ is the attention map from the last fixation at time t_0 and ϕ represents the empirical density map[2]. The exponential function indicates that with increasing fixation duration $(t - t_0)$ the influence of the previous attention map $A(t_0)$ decreases. The parameter ω controls the speed of the updating process for the map. The fixation map F at time t is computed as

$$F(t) = \frac{G_F}{\sum G_F} + e^{-\omega_f(t-t_0)} \left(F(t_0) - \frac{G_F}{\sum G_F} \right), \tag{1.3}$$

where G_F is equivalent to G_A from the attention map, but with different standard deviation σ_f. The fixation map updates like the attention map, but the empirical density does not play a role in creating this map. After the attention map and the fixation map at time t are computed, they are combined to the target selection map u in the following way:

$$u_{ij}(t) = \frac{[A_{ij}(t)]^\lambda}{\sum_{kl}[A_{kl}(t)]^\lambda} - c_{inhib} \frac{[F_{ij}(t)]^\gamma}{\sum_{kl}[F_{kl}(t)]^\gamma} \ . \tag{1.4}$$

In this equation c_{inhib} is a constant parameter controlling the strength of the inhibition, which was not present in the original model version (Engbert et al., 2015) but was added in a newer version (Schütt et al., 2017). The parameters λ and γ control the variance within the fixation and attention map and λ is generally set to 1. Since u is a target selection map, where each value should translate to a probability for being fixated, negative values have to be replaced. Thus, each value of u which is smaller or equal to zero is set to zero on target map u^* as

[2]The empirical density map can only be used when data from eye tracking experiments for an image are available. To produce scanpaths on any given input image, any fixation location or saliency model can be used instead of the empirical fixation map ϕ.

$$u_{ij}^*(t) = \begin{cases} u, u \geq 0 \\ 0, u < 0 \ . \end{cases} \tag{1.5}$$

Since this still produces impossible target locations, which is not plausible, due to noise and measuring error, the adjusted u^* is further weighted by factor η, to obtain a probability map $\pi(i, j)$ for each field on the grid. This is computed as

$$\pi(i,j) = (1-\eta)\frac{u_{ij}^*}{\sum\limits_{kl} u_{kl}^*} + \eta\frac{1}{\sum\limits_{kl} 1}, \tag{1.6}$$

to obtain small positive values for each location which was zero on u_*. This weighting function does not distort locations with high activities and has a sum of 1, making it a probability distribution. The step of removing values which equal zero from the map is crucial for estimating the likelihood of the model. If any fixation were to fall onto a grid cell with zero probability, the likelihood of the model to have produced the whole scanpath would be zero (see next section). In the original version of the model (Engbert et al., 2015) all vallues of u, which were smaller than η, were set to η.

Note, that activations from the fixation map F have a negative influence on u, whereas parameters from A have positive influences. In the map u, a higher value means a higher probability for receiving a fixation. Thus, locations with high activations on fixation map F and low activations on attention map A have a small probability for being fixated on the next fixation, meaning that the fixation map inhibits locations close to the current fixation and the attention map excites locations close to the current fixation. This might seem odd at first, because the maps are computed rather similarly but work in a reverse direction. However, the width of the standard deviation of the Gaussians (σ_f and σ_a) are different and the paramaters ω_a and ω_f controlling the dynamic of the maps are different as well. In the original model version by Engbert et al. (2015) and a in a newer version by Schütt et al. (2017), the attention map's standard deviation was larger than the fixation map's standard deviation. Thus, inhibition of return is more local than the allocation of visual attention around the fixation position. In both model versions the decay parameter for the attention map is larger than that of the fixation map. This means that inhibition is slower and lasts longer than the influence of early visual processing. New results have shown that a divisive instead of a subtractive influence of the fixation map F on the target selection map u might improve results (Schütt et al., 2017), however, in the studies of this thesis, we simulated data with the subtractive version of the model. In Chapter 2 we used to original version of the model (Engbert et al., 2015) and in Chapter 3 the newer version (Schütt et al., 2017).

After obtaining a probability map $\pi(i, j)$ for each possible fixation location, a fixation

is drawn according to the probability on the map π. This means that the creation of the map π itself is deterministic for each scanpath history, but fixation selection follows a stochastic process.

Engbert et al. (2015) have shown that the SceneWalk model is able to replicate typical scanpath statistics like saccade amplitudes and the pair-correlation function of fixations, which describes how strong fixations within one observation tend to cluster (Engbert et al., 2015; Trukenbrod et al., 2017). Additionally, we have shown that the model substantially outperforms sampling from the empirical density map, which by definition is the perfect static saliency map (Schütt et al., 2017). Thus, the model has already proven to capture some dynamic aspects of a scanpath. The studies within this thesis were designed to further elaborate the model and evaluate the model regarding dynamical aspects of a scanpath, which had not yet been tested.

1.4.5 Evaluating dynamical models of scanpath generation

The development of dynamical models led to new approaches in scanpath modeling. Unfortunately, it turned out that it is very difficult to compare dynamical models in terms of performance metrics. Le Meur and Baccino (2013) reviewed commonly used methods for scanpath comparison. One limitation of many evaluation methods is that they require predefined areas of interest within an image. Additionally, all methods described capture different aspects of scanpaths and are thus difficult to compare. To overcome this obstacle, we developed a likelihood-based method which facilitates the comparison of dynamical models and additionally provides an efficient way to estimate optimal model parameters (Schütt et al., 2017). The likelihood of any dynamical model can be computed if the model output is a gridded target map with activations, which can be transformed into probabilities for receiving a fixation. If a model is deterministic in the creation of the target selection map, the likelihood of the model can be computed rather effortlessly.

For each fixation of a scanpath, the SceneWalk model produces activations at each possible target location as a function of viewing time. To obtain the likelihood of a scanpath, all likelihood values of the empirically observed fixations are multiplied. This likelihood value can then for example be compared to other dynamical models, to a random fixation selection or to an image independent central fixation bias (Schütt et al., 2017). For better interpretation, the likelihood is usually logarithmized to the base of 2. The difference of the log-likelihood between two models is called the information gain of the better performing model compared to the other one. The advantage of this evaluation method is that it captures not only the fixation location positions, but all aspects of the scanpath.

We used the likelihood approach for the evaluation and parameter estimation of an

extended SceneWalk model in Chapter 3. The likelihood method also showed that the SceneWalk model outperforms sampling from the empirical density by far, which confirms that knowing all fixation locations is by itself not enough to predict valid scanpaths. Comparisons between the SceneWalk model and other dynamical models of saccade generation are subject to future research and will reveal weaknesses and advantages of the different dynamical models.

1.5 The present studies

Chapters 2–4 of this thesis consist of three studies investigating the contribution of selected systematic eye-movement tendencies on scanpaths in natural scene viewing and natural scene search. All studies were conducted to evaluate and improve the SceneWalk model of scanpath generation.

Chapter 2 will report data from a memory experiment, Chapter 3 from four free-viewing experiments and Chapter 4 from a visual search experiment. The studies report eye-movement recordings from 132 participants who saw between 25 and 120 images. Overall, fixations and saccades from about 25,000 experimental trials are investigated in this thesis.

1.5.1 Influence of initial fixation position in scene viewing

The experiment reported in Chapter 2 was designed to improve our understanding of the dynamic interaction between the location of the first fixation of a scanpath, image saliency, and the evolution of the scanpath. For the SceneWalk model this study serves as a validation of the spatial inhibition of return mechanism incorporated by the fixation map. Results from this study demonstrate the importance of dynamic aspects for computational models of scanpath generation.

In the large body of literature investigating eye movements in natural scene viewing, the starting position of the eyes has almost entirely been neglected. In most experiments participants start their observation in the center of the screen (see Tatler, 2007) and, after the scene is presented, participants are allowed to move their eyes. This first fixation, close to the center of the screen, is then removed from further analysis, because it's location was induced by the experimental design.

It has been shown that the first glance of a scene provides the observer with a relatively good representation of a scene, often described as the scene's gist (for an overview on gist see Oliva, 2005). Thus, it seems to be a valid assumption that the initial fixation of a scanpath plays an important role in how humans further explore a scene.

For our first experiment we manipulated the initial fixation position in a scene-viewing

experiment. Participants were forced to maintain fixation for one second after appearance of the image on the starting position, which was close to the left or right border of the screen. After participants maintained fixation for a second, the fixation cross disappeared and they were instructed to inspect the image for 10 s for a successive memory test.

Results showed a strong interaction between visual saliency and starting position and an influence of the starting position on the scanpath for up to 5 s. Saliency distribution was measured as a combination of two well known saliency models (Harel et al., 2007; Judd et al., 2009) and the empirical distribution of fixations. A strong overshoot to the image side opposite of the starting position was observed in 7 of 8 conditions and lasted for up to 5 s. This longlasting influence of the starting position is remarkable, since most scene-viewing studies (i) do not take the starting position into account and (ii) only last for around 3-5 s (Bylinskii et al., 2015). This study additionally revealed an asymmetry between left and right starting positions. If participants started on the right image side, the first saccade was significantly larger and overshoots to the other image side were stronger. These results agree with previously found leftward biases in natural scene viewing (Dickinson & Intraub, 2009; Ossandón et al., 2014).

For the SceneWalk model, this study was conducted to gain information about the interaction of early fixation positions and stimulus material. Additionally, we conducted this experiment to validate a basic model component, the inhibitory fixation map. For this purpose we simulated data with the SceneWalk model, a selection of statistical models and a model which incorporates the distribution of successive saccadic amplitudes and angles without an inhibitory component. We used the empirically observed starting positions, fixation durations and number of fixations for each trial to start simulations[3]. Only models with an inhibitory component were able to reproduce the observed overshoot, the model based on the angular distribution was not able to recreate this particular scanpath characteristic and neither was sampling from the empirical density map. Saccades returning immediately back to the starting positions were hardly observed in this experiment.

The results from this study advocate the idea that an inhibition-of-return mechanism is a valid driving force for dynamical models of scene perception. It also confirms that knowing all fixation locations on an image is not enough to predict valid scanpaths. The long-lasting influence of the starting position proves that it is crucial to know the starting position, with which observers were first confronted with the stimulus, to adequately model scanpaths in natural scenes.

In summary, this study is the first to show that the initial fixation position has a long-lasting influence on further scanpath progression. The results and model simulations support spatial inhibition of return in natural scene viewing, which is a fundamental

[3]For the model, which incorporates the distribution of saccadic amplitudes and angles, we used the first two fixations because this model needs an initial saccade direction to compute further saccades

principle of the SceneWalk model of scanpath generation.

1.5.2 The temporal evolution of the central fixation bias in scene viewing

Chapter 3 reports a study designed to systematically investigate the issue of the central fixation bias (CFB) in scene viewing. The central fixation tendency is a strongly systematic eye-movement behavior in scene-viewing experiments (Tatler, 2007). This bias is extremely persistent and is particularly pronounced on the initial fixations of an observation.

The CFB has been investigated thoroughly in previous work (e.g., Buswell, 1935; Tatler, 2007; Bindemann, 2010), but the underlying cause has not yet been fully identified. The CFB has been found in many scene viewing experiments, regardless of the starting position (Tatler, 2007), the image position on the screen (Bindemann, 2010), the images' low-level features (Tatler, 2007) and many more. Due to the omnipresence of the central fixation bias and the fact that it is especially pronounced on early fixations, it possibly masks attentional allocation driven by top-down or bottom-up processes.

This thesis aims to obtain new knowledge about eye movement behavior in natural scene viewing to predict scanpaths and to understand the underlying cognitive processes which lead to these scanpaths. Since the CFB is the best single predictor for fixation locations in natural scene viewing (Judd et al., 2009; Vincent et al., 2009; Clarke & Tatler, 2014; Kümmerer et al., 2015; Schütt et al., 2018), it is important to find out if the CFB is a laboratory artifact or whether it can be transferred to natural viewing behavior and, if so, why it exists.

The manipulation of our experiment from Chapter 2 led to a reduction of the CFB compared to scene-viewing experiments, in which the initial fixation position was not experimentally prolonged. To further investigate the cause of this reduction, we conducted four experiments with variable starting position and an experimentally prolonged initial fixation by variable time intervals from 0-1 s. This was done by presenting an image with a pretrial fixation cross present for a certain amount of time and instructing participants not to start exploration until the fixation cross had disappeared. Our hypothesis was that this manipulation dissociates the sudden image onset and the signal to move the eyes, which we hypothesized would reduce the image independent CFB. We used the same images as Tatler (2007) in his seminal work on the CFB.

The results confirmed our observations from Chapter 2. A delay of the initial saccade led to a reduction of the CFB, if this delay was equal to or larger than 75 ms. Smaller delays did not reduce the CFB and increasing the delay above 250 ms did not produce effects noteworthy. Analyzing the initial saccade latency, regardless of our manipulation,

showed that this initial latency reliably predicts the magnitude of the CFB. Short initial latencies led to a strong CFB and longer latencies led to less CFB. This confirms the idea that the CFB is mainly an image independent artifact, which is produced by the sudden onset of an unknown stimulus. The results from this study were especially helpful for implementing a biologically plausible CFB into the SceneWalk model. For the initial attention map of the SceneWalk model, we used a map with a central activation instead of the Gaussian weighted empirical saliency map from the original model. Thus, with increasing time, this central activation map transforms into the fixation location dependent map from the original model. The adjustment replicated the latency-dependent CFB in a plausible and computationally rather simple matter. The adjusted model reproduced the qualitative progression of the CFB throughout the trial and improved model performance on initial fixation selection substantially. The design of the experiments from Chapter 3 reduced the influence of the sudden image onset and thus might create a more natural viewing experience for the participants.

In summary, we found a way to reliably reduce the central fixation bias in this study, which mainly depends on the initial saccade latency. We used our results to implement a plausible central fixation bias in the SceneWalk model.

1.5.3 Searchers adjust their eye-movement dynamics according to the search target in natural scenes

The last study of this thesis, presented in Chapter 4, contains a visual search experiment. In visual search on natural scenes, fixation locations are influenced by the visual properties of the search target (e.g., Wolfe, 1994; Wolfe & Horowitz, 2004; Hwang et al., 2009). To our knowledge it has not been investigated whether the target's visual properties also influence other scanpath properties like fixation durations, saccade amplitudes or changes in saccadic direction. Answering the question, whether humans adapt their search behavior to the target is crucial to model scanpaths in visual search experiments, since dynamical parameters would depend on the search target, if participants actually adjusted their eye movement characteristics to the target.

Many visual search experiments have been conducted on so called search arrays, where one target and multiple distractors are presented on homogeneous background and the goal is to find the target. Although these studies have provided many new insights into how visual search works, they mostly disregard the role of eye movements, which are an important aspect of many visual search tasks (Zelinsky, Rao, Hayhoe, & Ballard, 1997; Findlay & Gilchrist, 1998, 2003; Rayner, 2009; Hulleman & Olivers, 2015).

Eye movements in complex visual search have shown that human searchers not simply fixate the most likely target location but move their eyes almost optimally such that the

probability of finding the target, according to properties of our visual field (the degradation of visual acuity into the periphery), is maximized (Najemnik & Geisler, 2005, 2008; Geisler, 2011). Because target features influence fixation locations on complex backgrounds so strongly, we were interested in how they influence other basic scanpath properties like saccade amplitudes and fixation durations.

In the study we showed participants 6 artificial targets of different spatial frequency content, to find out whether scanpath properties are adjusted to the visibility of the targets in the periphery. Eye movements were adjusted to the visual properties of the target efficiently. High-spatial frequency targets, which are less visible in the periphery, led to smaller saccade amplitudes than low-spatial frequency targets. Fixation durations were also adjusted to the target, such that high-spatial frequency targets led to shorter fixation durations. High-spatial frequencies can be perceived better in central vision (Laubrock et al., 2013; Schütt & Wichmann, 2017), thus it is useful to make more eye movements with shorter fixation durations when looking for high-spatial frequency targets. A recent model of early spatial vision (Schütt & Wichmann, 2017) evaluated the targets in terms of foveal detectability. This evaluation showed that the high-spatial frequency targets had a higher signal to noise ratio compared to the background when presented in the central visual field. This additionally indicated that short fixation durations are more useful when looking for high-spatial frequency targets.

In a post-hoc analysis we found that only saccades which changed direction compared to the previous saccade were influenced by target properties. Saccades which maintained direction from the previous saccade did not differ between the two target types. The absence of a target influence brought us to the idea that saccades which maintain direction are part of a default scanning mechanism. Analyzing saccades without a change in direction in terms of visual saliency and empirical density confirmed this assumption, because they landed on positions which were less salient and less looked at by other observers compared to other fixation locations.

In summary, this study is the first to show that human participants adjust their fixation durations and saccade amplitudes to the visual features of the target in complex visual search. Additionally, the results from this study provide new information about the role of forward and backward saccades in natural scene search.

Chapter 2

Influence of initial fixation position in scene viewing

Lars O. M. Rothkegel[1*], Hans A. Trukenbrod[1], Heiko H. Schütt[1,2], Felix A. Wichmann[2–4], and Ralf Engbert[1]

[1]University of Potsdam, Germany
[2]Eberhard Karls University Tübingen, Germany
[3]Bernstein Center for Computational Neuroscience Tübingen, Germany
[4]Max Planck Institute for Intelligent Systems, Tübingen, Germany

Running head: Initial fixation position

published 2016 in *Vision Research, 129,* 33–49.
doi: 10.1016/j.visres.2016.09.012

Abstract

During scene perception our eyes generate complex sequences of fixations. Predictors of fixation locations are bottom-up factors such as luminance contrast, top-down factors like viewing instruction, and systematic biases e.g. the tendency to place fixations near the center of an image. However, comparatively little is known about the dynamics of scanpaths after experimental manipulation of specific fixation locations. Here we investigate the influence of initial fixation position on subsequent eye-movement behavior on an image. We presented 64 colored photographs to participants who started their scanpaths from one of two experimentally controlled positions in the right or left part of an image. Additionally, we used computational models to predict the images' fixation locations and classified them as balanced images or images with high conspicuity on either the left or right side of a picture. The manipulation of the starting position influenced viewing behaviour for several seconds and produced a tendency to overshoot to the image side opposite to the starting position. Possible mechanisms for the generation of this overshoot were investigated using numerical simulations of statistical and dynamical models. Our model comparisons show that inhibitory tagging is a viable mechanism for dynamical planning of scanpaths.

2.1 Introduction

An important problem for research on human vision is to predict where people look in visual scenes (Tatler & Vincent, 2008). Recording of eye movements is among the most important tools to investigate how attention is distributed over a given scene (Findlay & Gilchrist, 2003). In addition to scene content (Henderson, 2003), image-independent viewing strategies exist, e.g., the central fixation tendency (Tatler, 2007) as the most important effect in this category. To obtain a deeper understanding about dynamical aspects of the attention distribution over a scene and possible dependencies between successive fixations we investigate the influence of the eye's starting position on subsequent viewing behavior based on statistical and dynamical assumptions about eye guidance.

Processes that influence the selection of upcoming saccade targets can be divided into three different categories of theoretical principles. *Bottom-up processes* derive from properties of the viewed stimulus (S. K. Mannan, Ruddock, & Wooding, 1996; Itti et al., 1998; Parkhurst et al., 2002). *Top-down processes* depend on the mental state of an observer, e.g., the observers' visual memory (Henderson & Hollingworth, 2003) or the instruction given to the observer before inspection of a scene (Yarbus et al., 1967; Castelhano et al., 2009). Finally, *systematic tendencies* describe eye movement behavior found in many experiments independent of stimulus and observer. The initial selection of the center of an image (Tatler, 2007; Bindemann, 2010), the tendency to make initial movements in the leftward direction (Dickinson & Intraub, 2009; Foulsham, Gray, Nasiopoulos, & Kingstone, 2013; Ossandón et al., 2014) or the preference for horizontal and vertical over oblique saccades relative to the image (Foulsham & Kingstone, 2010) belong to this category.

Research on bottom-up processes has been particularly popular to predict fixation locations from low-level image features such as contrast, orientation and color (Itti et al., 1998; Torralba, 2003; Kienzle et al., 2006). For a given scene, computational models generate a *saliency map*, a 2D probability distribution that indicates the probability of receiving a fixation in an eye tracking experiment with human participants (Itti et al., 1998; Itti & Koch, 2000; Judd et al., 2009; Borji & Itti, 2013). Thus, a saliency map is a stationary model that computes probabilities for all locations simultaneously.

However, current computational models for the prediction of fixation locations are not exclusively based on bottom-up features. Recent models incorporate top-down processes like task demands (Navalpakkam & Itti, 2005) and other higher-level image features like face processing (Cerf et al., 2008). Moreover, systematic tendencies such as the central fixation bias (Tatler, 2007) are included in the computation of fixation density models. As a result, current models integrate multiple features from all three categories of processes into a coherent computational framework (Cerf et al., 2008; Judd et al., 2009; Kümmerer

et al., 2015). Although the original meaning of *saliency* refers to the bottom-up features of an image, newer computational models that include other features are also termed saliency models by their authors (Judd et al., 2009; Bylinskii et al., 2015). Because of this unclear terminology we will refer to all stationary models that aim at the prediction of fixation locations as fixation density models. A location that a model tags as likely to receive a fixation will be reffered to as conspicuous rather than salient.

All fixation density models need to predict the density of the eye's fixation locations (so-called first-order statistics). Thus, the evaluation of the models is primarily based on the assumption of statistically independent fixations without reference to previous fixations, i.e., the scanpath (Kümmerer et al., 2015). In contrast to static models, dynamic models try to capture some additional aspects of the scanpath. Dynamical principles for saccade planning are *inhibitory tagging* (Klein, 1988; Itti et al., 1998; Bays & Husain, 2012; Le Meur & Liu, 2015), *saccadic momentum* (Smith & Henderson, 2009, 2011; Wilming et al., 2013) and *facilitation of return* (Smith & Henderson, 2009, 2011; Luke, Schmidt, & Henderson, 2013)

Inhibitory tagging is motivated by the effect of inhibition of return, a neural mechanism that inhibits the processing at recently attended locations (Posner & Cohen, 1984; Posner et al., 1985; Klein, 2000) and is often interpreted as a foraging facilitator. While this mechanism was first discovered as an effect on a temporal scale, i.e., increased processing time at a previously attended stimulus for a specific time window, inhibition of return might carry over to spatial effects. In the case of spatial inhibition of return recently fixated positions are inhibited from being re-fixated shortly afterwards (Gilchrist & Harvey, 2000). Several studies were unable to report evidence for inhibition of return during scene viewing; quite the contrary, a facilitation of return saccades to currently fixated locations has been found (Smith & Henderson, 2009, 2011; Wilming et al., 2013).

However, compared to a statistical baseline model without memory based on inhibitory tagging, return saccades occur less often in experiments than expected (Bays & Husain, 2012), when the density map of fixations and the distribution of angles between two subsequent saccades are reproduced. Therefore, there is at least weak support for a memory-producing mechanism during scene exploration. In agreement with this result, we recently published a computational model of saccade generation in scene viewing that implements both inhibitory tagging and dynamical attention mechanisms (Engbert et al., 2015). In this model inhibitory tagging is combined with a dynamical activation map representing attention allocation, allowing the model to reproduce second-order statistics that include spatial correlation functions characterizing the clustering of fixations in addition to the first-order density of fixations. Thus, inhibitory tagging seems to be important to reproduce higher-order scanpath statistics (Engbert et al., 2015), despite the current lack of direct experimental support for inhibition of return in scene viewing (Smith &

Henderson, 2009, 2011; Luke, Schmidt, & Henderson, 2013).

Saccadic momentum, another dynamical principle of saccade planning in scene viewing, describes the tendency to maintain the direction of the previous saccade for the upcoming saccade (Smith & Henderson, 2009, 2011; Wilming et al., 2013). Similar to inhibition of return, saccadic momentum could serve as a foraging facilitator in visual search. Finally, *facilitation of return* describes the tendency that it is actually more likely to produce return saccades than it would be by chance (Hooge, Over, van Wezel, & Frens, 2005; Smith & Henderson, 2009). On the time scale of one fixation duration ($\sim$ 300 ms), such a facilitation seems to be in contradiction to spatial inhibitory tagging. Because of these behaviorally relevant processes, we were interested to find experimental support for the presence of *inhibitory tagging*, *saccadic momentum*, *facilitation of return* or a mixture of these fundamental principles in attentional and oculomotor control.

Smith and Henderson (2009) ruled out inhibitory tagging, since they found an increased number of return saccades in comparison to a probabilistic baseline (Smith & Henderson, 2009). However, it has also been argued that there is a reduced number of return saccades compared to a memoryless system (Bays & Husain, 2012). Given the current mixed evidence on return saccades, we focus on the time window of events. Return saccades are limited to a time window of one fixation duration, i.e., about 300 ms. Since attention moves to the future fixation location before a saccade is executed (Deubel & Schneider, 1996), inhibition of return is at its maximum shortly after the saccade is planned if we assume that the typical time-course transfers to scene viewing (Posner & Cohen, 1984; Klein, 2000). However, first, it would not be surprising to find that more time than a single fixation duration is needed to build-up spatial inhibition. Second, return saccades might be planned before the inhibition of return mechanism is activated, so that saccades to previously inspected image regions could be produced while inhibition is on the rise. Third, it has been reported that the time scale of IOR is dependent on task difficulty (Klein, 2000). Therefore, the current lack of direct evidence for inhibition of return does not rule out inhibitory tagging as a saccade-planning mechanism.

To investigate inhibitory tagging, saccadic momentum, and facilitation of return, we recorded observers' scanpaths on natural scenes starting from one of two predefined starting positions close to either side of the monitor. Participants were forced to maintain fixation at an initial location in an image for one second under gaze-contingent monitoring. Under the hypothesis that spatial inhibitory tagging is active at the starting position, we expected observers (i) to leave their starting positions when fixation markers disappeared, and (ii) not to return immediately to the region of the experimentally controlled starting position. Since we hypothesized that both behaviors depend on the conspicuity of the region of the starting position, we classified natural images into three categories with left-sided and right-sided conspicuity asymmetry as well as images with an approx-

imately symmetrical distribution. First, we expected that initial fixations stay closer to the starting position when the starting position was in interesting side of a scene; second, gaze was expected to move immediately to the opposite side of a scene, when the starting position was opposite to the scenes interesting side. Third, according to the saccadic momentum and facilitation of return hypothesis, we expected a behavior where subsequent eye movements depend on the direction of the first saccade. With the typical center bias we assume that the gaze had to shift to the center and, subsequently, either maintain direction and move to the opposite image side (saccadic momentum) or return close to the starting position (facilitation of return).

Below we report that gaze positions of the participants moved further away from the starting position than predicted by the empirical fixation map or a saccadic momentum mechanism. Next, we compare experimental data with numerical simulations from a range of models, including a model reproducing the saccadic momentum mechanism and our dynamical model (SceneWalk) which uses inhibitory tagging as a mechanism for saccade planning (Engbert et al., 2015) and a combination of the latter ones.

2.2 Method

The methodology of this work is similar to a recently published study from our lab (Engbert et al., 2015).

2.2.1 Experiment

Stimuli

A set of 64 color photographs was presented to human observers. Pictures were presented on a 20 inch CRT monitor (Mitsubishi Diamond Pro 2070; frame rate 120 Hz, resolution 1280 × 1024 pixels; Mitsubishi Electric Corporation, Tokyo, Japan). The dimensions of the monitor where 39.6 cm (horizontal) x 29.7 cm (vertical) and the viewing distance was 70 cm. For the presentation during the experiment all images were converted to a size of 1200 × 960 pixels and displayed in the center of the screen with gray borders extending 32 pixels to the top/bottom and 40 pixels to the left/right of the image. Images covered 31° of visual angle in the horizontal and 25° in the vertical dimension.

Images showed either natural object-based scenes (N=48) or abstract natural patterns (N=16). All photographs were taken by members of our lab. Object-based scenes were further devided into three categories as balanced, left focus, or right focus, yielding a total of 4 categories (Fig. 2.1). The Pattern images were chosen to obtain a more homogenous fixation distribution because of the lack of objects present. Systematic oculomotor biases were expected to be more evident in these images.

For the categorization of object-based scenes we used an objective test by computing conspicuity with the graph based visual saliency model (Harel et al., 2007) and the Judd model (Judd et al., 2009) without distance to center weighting and face or object detection. As a posthoc measure, the density map of the observers' fixations for each of the 48 natural scenes was evaluated to obtain an empirical measure of left and right bias for the images. Figure 2.2 shows an example of an image with right focus compared to the output of the two computational models and the kernel density estimate of the fixation density (excluding the initial fixation) from all observers. To obtain a quantitative measure for the presence of a left or right focus, we computed the horizontal position of a vertical line with equal conspicuity/intensity on each side. If the horizontal position of this line differed by more than 5 percent from the center (for the average over the two computational models and the human fixation map), the corresponding image was classified as having left or right focus. After application of this criterion, we retained 23 images with focus close to the center (balanced images), 12 images with left focus, and 13 images with right focus among the set of object-based scenes[1]. The distribution of focus for the different models as well as the rater's judgements are shown in Figure 2.3. Though for some images the focus of the empirical fixation map differs strongly from the computational models, overall they match fairly well. The green line, that represents the empirical map lies below the other lines. This indicates that human fixation locations are biased more to the left than the computational models predict. This is compatible with the findings of an initial leftward bias in scene viewing (Dickinson & Intraub, 2009; Foulsham et al., 2013; Ossandón et al., 2014).

Participants

We recorded eye movements from 28 human participants with normal or corrected-to-normal vision. The group of participants consisted of 20 female and 8 male observers aged between 19 and 33 years; all were recruited from the University of Potsdam. Participants received credit points or a monetary compensation of 8 Euro for their participation. The work was carried out in accordance with the Declaration of Helsinki. Informed consent was obtained for experimentation by all participants.

Procedure

Participants were instructed to position their heads on a chin rest in front of a computer screen. Eye movements were recorded binocularly using an Eyelink 1000 video-based eye tracking system (SR Research, Osgoode/ON, Canada) with a sampling rate of 1000 Hz.

[1]Based on our subjective assesment each category contained 16 images. Because our subjective categorization did not match the objective criterion for some of the images, an unequal number of images in each category remained for further analysis.

Figure 2.1: Examples from the set of images, (a) balanced (b) natural pattern (c) left focus (d) right focus.

Trials began with a black fixation cross presented on a grey background at the vertical meridian 5.6° away from the left or right border of the monitor. After successful binocular fixation in a square with a side length of 2.2° an image appeared while the fixation cross remained present for another second. Participants were instructed to keep their eyes on the fixation cross until it disappeared. This was done to assure that participants started their exploration from the experimentally controlled position. If this fixation test failed, a mask with random noise appeared and the fixation check was repeated. After successful completion of the fixation test participants explored each scene for 10 s for a subsequent memory test. In the memory test participants had to indicate for 64 images—32 presented images and 32 new images—if they had seen it before.[2] Figure 2.4 summarises the experimental procedure. In the example, the first fixation test failed, before the actual scene exploration started. A fixation check of 1 second turned out to be very difficult for participants and had to be repeated in 32% of all trials. Thus, some participants experienced an even longer preview from the starting position before the actual trial. Importantly, no participant was able to fixate the image from a different

[2]Participants answered correctly in 91.5% of all trials with a mean reaction time of 1.4 seconds.

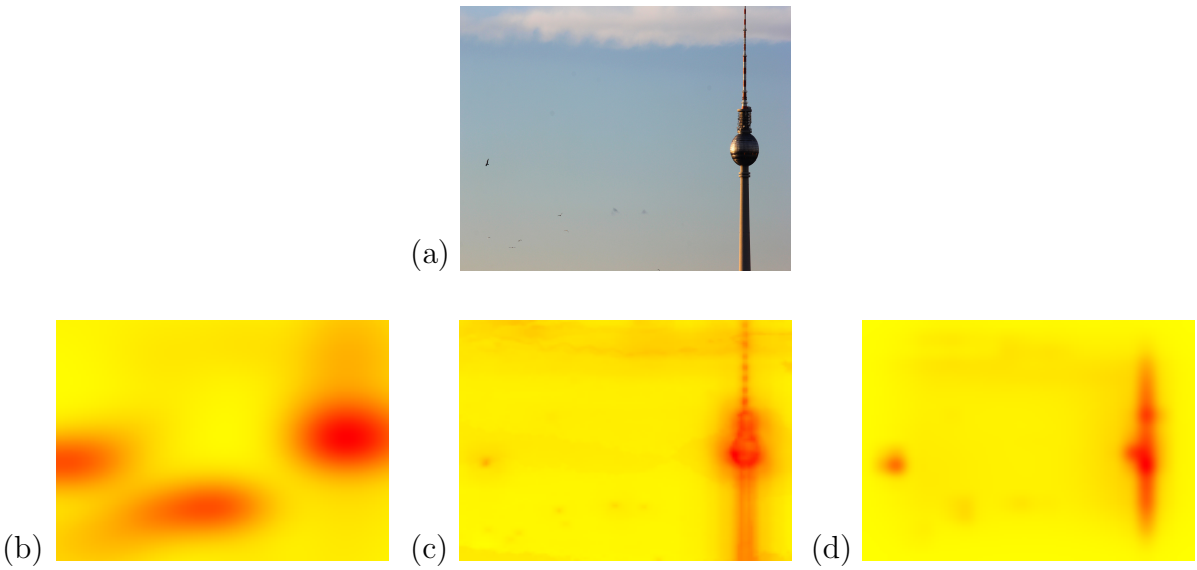

Figure 2.2: Objective categorizing of images. (a) Example of an image with right focus. (b) Experimental density map of fixations, estimated using a Gaussian kernel with bandwith $\sigma = 2.56°$ according to Scott's rule. (c) Output from the Judd model, without distance-to-center analysis and face/object detection. (d) Output from the graph based visual saliency (GBVS) model.

position before inspection and the fixation on the starting position was never shorter than one second. In 20% it was repeated once, in 6% twice and in 2.8% three times. In 2.2% percent of the trials the fixation test had to be repeated more than 3 times. All analyses were conducted separately for the trials with and without a repetition of the second fixation check. No systematic differences are visible between these analyses. The corresponding figures for the data without a repeated fixation check are provided as supplemantary material. For the analyses in the main text of this article we used fixations from all trials.

2.2.2 Data analysis

Data preprocessing and saccade detection

For saccade detection we applied a velocity-based algorithm (Engbert & Kliegl, 2003a; Engbert & Mergenthaler, 2006). This algorithm marks an event as a saccade if it has a minimum amplitude of 0.5° and exceedes the average velocity during a trial by 6 median-based standard deviations for at least 6 data samples (6 ms). The epoch between two subsequent saccades is defined as a fixation. The number of fixations for further analyses was 47 330.

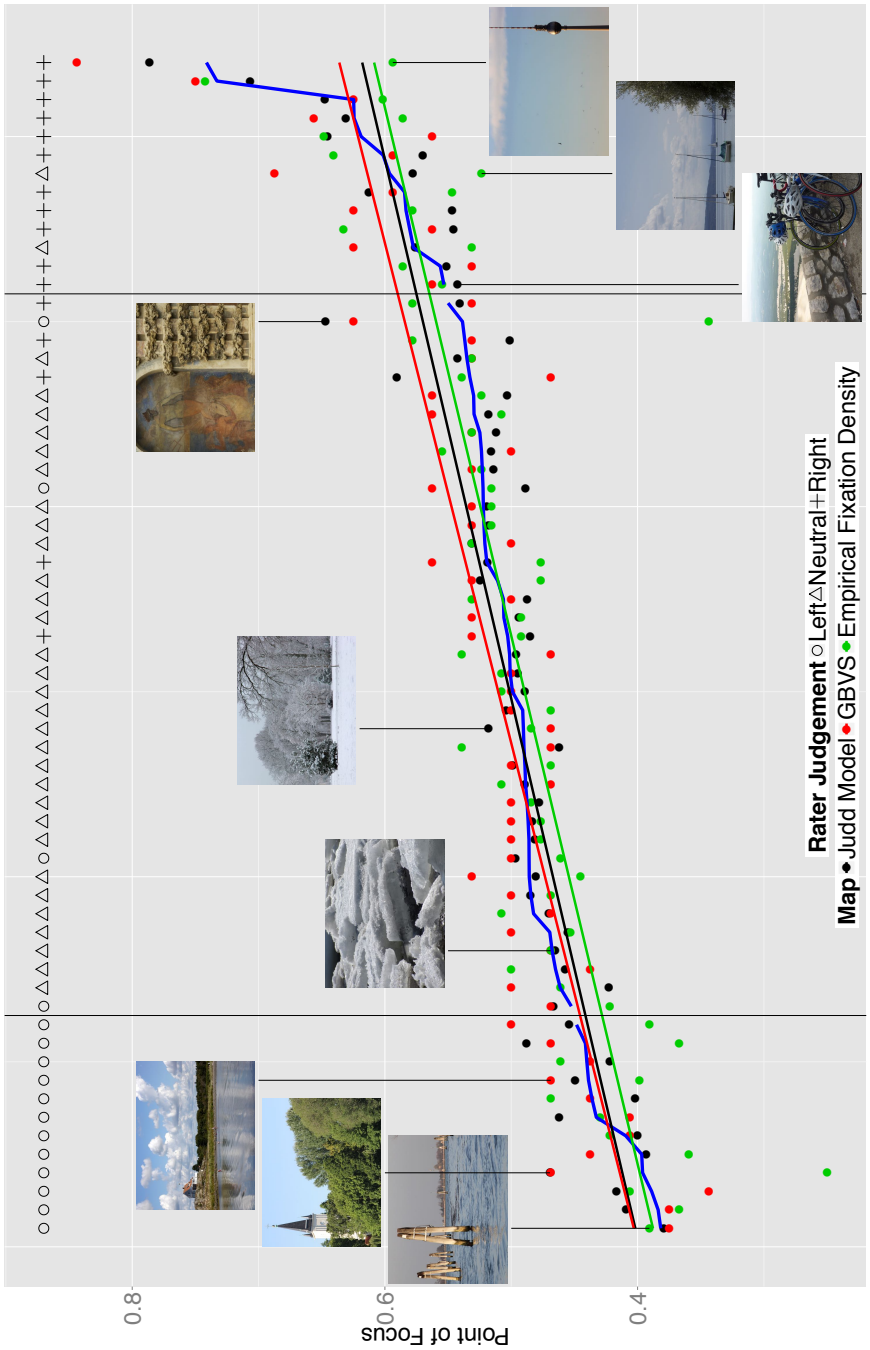

Figure 2.3: Categoriziation of images. All images ordered by point of focus from strongest left to strongest right focus image. The blue line is the mean value of the three categorization measures. The black vertical lines indicate where the posthoc measure devided between left focus, neutral and right focus images. Symbols at the top of the graphic show the rater's judgements of the images.

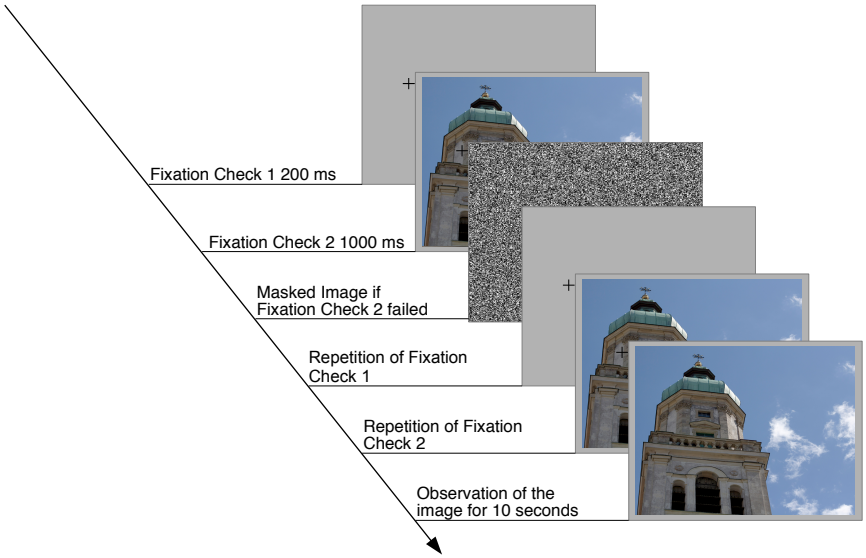

Figure 2.4: Schematic illustration of the experimental procedure. In the example, the fixation test failed. After a repetition of the fixation check exploration of the image started.

Mean horizontal distance from starting position

To analyse the potential dependence of the scanpath on the experimentally controlled starting position, we estimated the temporal evolution of the mean horizontal gaze position. In the first step, we computed the time-dependent horizontal distance to the starting position for each trial. The calculation was based on fixation positions and fixation durations obtained from data preprocessing. The estimated mean horizontal distance (MHD) from the starting position was computed as

$$X_{\mathrm{MHD}}(t) = \frac{1}{m \cdot n} \sum_{j=1}^{n} \sum_{k=1}^{m} (x_{jk}(t) - x_{jk}(0)) \ ,$$

where $x_{jk}(t)$ indicates horizontal gaze position at time step t (in milliseconds) for participant j and image k. For each combination of image and participant, the starting position $x_{jk}(0)$ was close to the position of the initial fixation cross on the left or right side of an image (see Procedure). To obtain a comparable measurement for both starting positions, gaze position for right starters was mirrored on the vertical meridian. Afterwards a Gaussian kernel with $\sigma = 100$ ms was applied to obtain a smoothed curve of $\bar{X}_{\mathrm{MHD}}(t)$. Another possible analysis would be the vertical or the overall distance to the starting position. The vertical distance showed no interesting effect, as the starting position was always on the vertical midline. The overall distance did hence only depend on the horizontal distance, which we therefore analysed.

2.2.3 Model simulations with controlled initial positions

To interpret the experimental results of the temporal evolution of mean horizontal distance $X_{\mathrm{MHD}}(t)$ we performed numerical simulations using statistical control models, a model emulating saccadic momentum, a recently proposed dynamical model for scanpath generation using inhibitory tagging (Engbert et al., 2015) and a combination of the latter ones. For the model runs, simulations started at initial positions corresponding to the experimentally manipulated starting positions. Fixation durations and number of fixations in each trial were taken from the experimental data. We obtained the same number of trials from numerical simulations as from the experimental data and analysed the MHD function $X_{\mathrm{MHD}}(t)$ for each model. The number of grid points on which all models were computed was 128 in both dimensions.

Sampling from density map

As the most straightforward statistical control, we simulated scanpaths by randomly sampling from the 2D density map of all fixations on a given image, i.e., the empirical fixation map, generated by all participants. First, we applied kernel density estimation using the SpatStat package (Baddeley & Turner, n.d.) of the R Language for Statistical Computing (R Core Team, 2014). Based on a Gaussian kernel function with a bandwidth parameter according to Scott's rule (Scott, 2015), ranging from 1.81° to 2.72°, we computed the empirical fixation density map for each image. Second, to simulate a scanpath (i.e., a fixation sequence), we sampled randomly from this map where local density at a particular location translated into probability to generate a fixation at this position.

Gaussian Model

Next, we implemented a statistical model that sampled from the empirical fixation map via a Gaussian-shaped aperture to mimic a limited attentional span for saccade target selection. For a given fixation position x, the empirical fixation map was weighted by a two dimensional Gaussian, centered at x, with a standard deviation of 4.88° visual angle. The same standard deviation was used for the attention map of the SceneWalk model by Engbert et al. (see section 2.3.4). Sampling from the resulting weighted map, which was recomputed after each fixation, generated a scanpath in this model. Effectively, this model is similar to the SceneWalk model without an inhibitory tagging mechanism.

Saccadic Momentum Model

The third model reproduced the behavior that saccades, on average, tend to follow the direction of the previous saccade — a phenomenon termed saccadic momentum (Smith &

Henderson, 2009). In order to reproduce the typical angles between two subsequent saccades, while keeping the saccade amplitude distribution similar to the experimental data, saccades were sampled from the joint probability distribution of amplitudes and angles. This probability distribution was computed from all saccades with an amplitude smaller than 20°. We estimated a density map from these saccades with the SpatStat package (Baddeley & Turner, n.d.) where density translated into the probability of generating a saccade with length s and angle a. After sampling from this map, simulated saccades shifted fixation positions by length s and angle a with respect to the previous saccade. If this new target location did not lie within the image boundaries, a new saccade was sampled. To initialize the saccadic momentum model the first saccade and fixation position was taken from the experimental data.

SceneWalk model

In a recently proposed mathematical model for scanpath generation in scene viewing (Engbert et al., 2015), it was assumed that eye movements are driven by the interaction of two neural activation maps. A fixation map $f(x;t)$ keeps track of previous fixations by adding activation at fixation position x. The time dependence of this map results from the addition of activation at each time step in combination with fixation-position independent decay. The fixation map serves as an *inhibitory tagging* mechanism (Itti & Koch, 2001). The distribution of visual attention at time t is given by a second activation map $a(x;t)$. The assumption of maps of visual space is consistent with recent neurophysiological work on an allocentric motor map in the primate entorhinal cortex (Killian et al., 2012; Stensola et al., 2012), which is spatially discrete like that in the model with discrete activations $f_{ij}(t)$ and $a_{ij}(t)$, where subscripts i and j denote horizontal and vertical dimensions.

In the SceneWalk model, the difference of the normalized fixation map $f_{ij}(t)$ and the normalized attention map $a_{ij}(t)$ is a time-dependent potential function $u_{ij}(t)$ computed as

$$u_{ij}(t) = -\frac{a_{ij}(t)}{\sum\limits_{kl} a_{kl}(t)} + \frac{[f_{ij}(t)]^\gamma}{\sum\limits_{kl}[f_{kl}(t)]^\gamma} \,,$$

where the exponent γ is a free parameter that is important for controlling the amount of aggregation (or clustering) of realized gaze positions (Engbert et al., 2015).

Since the potential $u_{ij}(t)$ is the difference of activation maps, it can be positive or negative at position (i,j). We implemented stochastic selection of saccade targets proportional to relative activations (Luce, 1959) among the lattice sites with negative values

(S). The probability for saccadic target selection is given by

$$\pi_{ij}(t) = \max\left(\frac{u_{ij}(t)}{\sum\limits_{(k,l)\in\mathcal{S}} u_{kl}(t)},\, \eta\right) ,$$

where η is an additional model parameter that allows each grid position to serve as a possible saccade target with a probability above zero. All model parameters were chosen as in the published version of the SceneWalk model (Engbert et al., 2015). In an additional model run, the free parameter γ controlling the inhibition and amount of clustering of fixations was manually adapted for a second analysis for illustrative purposes.

SceneWalk Model + Saccadic Momentum

Because the original SceneWalk Model (Engbert et al., 2015) does not incorporate any information regarding the angular distribution between successive saccades we added a mechanism to the SceneWalk model to reproduce the distribution of angles. Before a saccade was chosen from the target map $u_{ij}(t)$ this target map was multiplied with a map representing the density function of angles with respect to the previous saccade. To obtain this map we first computed the angle that a saccade to each grid cell encloses with the previous saccade. Afterwards the probability of the angle was inserted into the grid cell. This probability was taken from a kernel density estimation of the angles between successive saccades. The resulting map was multiplied with the target map $u_{ij}(t)$ from the SceneWalk model and the combined map was normalized. This model thus behaved very similarly to the SceneWalk model but favoured grid cells that enclose empirically frequent angles between the previous and the future saccade.

2.3 Results

In our experiment, we manipulated starting positions to investigate the influence on scanpath statistics. We begin with reporting summary statistics on saccade amplitudes and saccade turning angles, before we analyze the temporal evolution of the mean horizontal distance from the starting position. The temporal evolution of the mean horizontal distance from the starting position will turn out to be an important measure of scanpath statistics. Finally, we run several numerical model simulations to interpret potential mechanisms underlying scanpath generation.

2.3.1 Saccadic amplitudes and directions

In our experiment, distributions of saccade amplitudes show the heavy tailed curve that is typically observed in scene viewing experiments (Tatler et al., 2006; Henderson & Hollingworth, 1998). Saccade amplitude distributions (Fig. 2.5a) across different image types and starting positions were very similar. The only visible difference was a slight shift from short to medium saccade lengths in the pattern images compared to the object based images.

We computed an ANOVA for the influence of saccade number on saccade amplitude. There was a significant effect between the first and the second saccade ($F(1, 1790) = 54.4, p = 2.5 \times 10^{-13}$), where the amplitude of the first saccade was larger than the amplitude of the second (Fig. 2.5b). Statistical tests between subsequent saccade amplitudes showed no significance, indicating that after the first long saccade, the mean amplitude reaches a stable value.

We computed another ANOVA to investigate influences of the image type and the starting position on the first saccade length. The starting position was significant ($F(1, 1785) = 47.95, p = 6.09 \times 10^{-12}$) as well as the image type ($F(3, 1784) = 11.73, p = 1.30 \times 10^{-7}$). The mean first saccade amplitudes for left and right starters were $\bar{s}_{\text{left}} = 7.80^\circ$ and $\bar{s}_{\text{right}} = 9.26^\circ$, resp. Mean values for the image types were $\bar{s}_{balanced} = 8.01^\circ$, $\bar{s}_{pattern} = 7.81^\circ$, $\bar{s}_{leftfocus} = 9.28^\circ$ and $\bar{s}_{rightfocus} = 9.02^\circ$. The interaction between image type and starting position was also significant ($F(3, 1784) = 53.67, p < 2 \times 10^{-16}$). Figure 2.5c visualizes this interaction and the main effects of image type and starting position.

In summary, forcing the observers to start exploration from an experimentally controlled initial position close to the border of the monitor resulted in a long first saccade. This was particularly true if the interesting image part was on the opposite side of the initial position. The longer initial saccade from right to left than vice versa is congruent to the left direction bias that has been found in various experiments (Dickinson & Intraub, 2009; Foulsham et al., 2013; Ossandón et al., 2014). This result indicates that the leftward bias is not only present, if participants start observations from the center of the image (see Fig. 2.3).

2.3.2 Saccade turning angle and its relation to amplitude

Statistically, most saccades are likely to follow the direction of previous saccades or shift gaze position back to the direction of the starting position of the previous saccade. The overall distribution of saccade turning angles between two subsequent saccades (Fig.2.6a) is characteristic for similar experiments in scene viewing (Tatler & Vincent, 2008; Smith & Henderson, 2009). Next, we constructed a conditional plot of saccade amplitude in relation to the previous saccade amplitude and orientation (Fig.2.6b). The endpoint

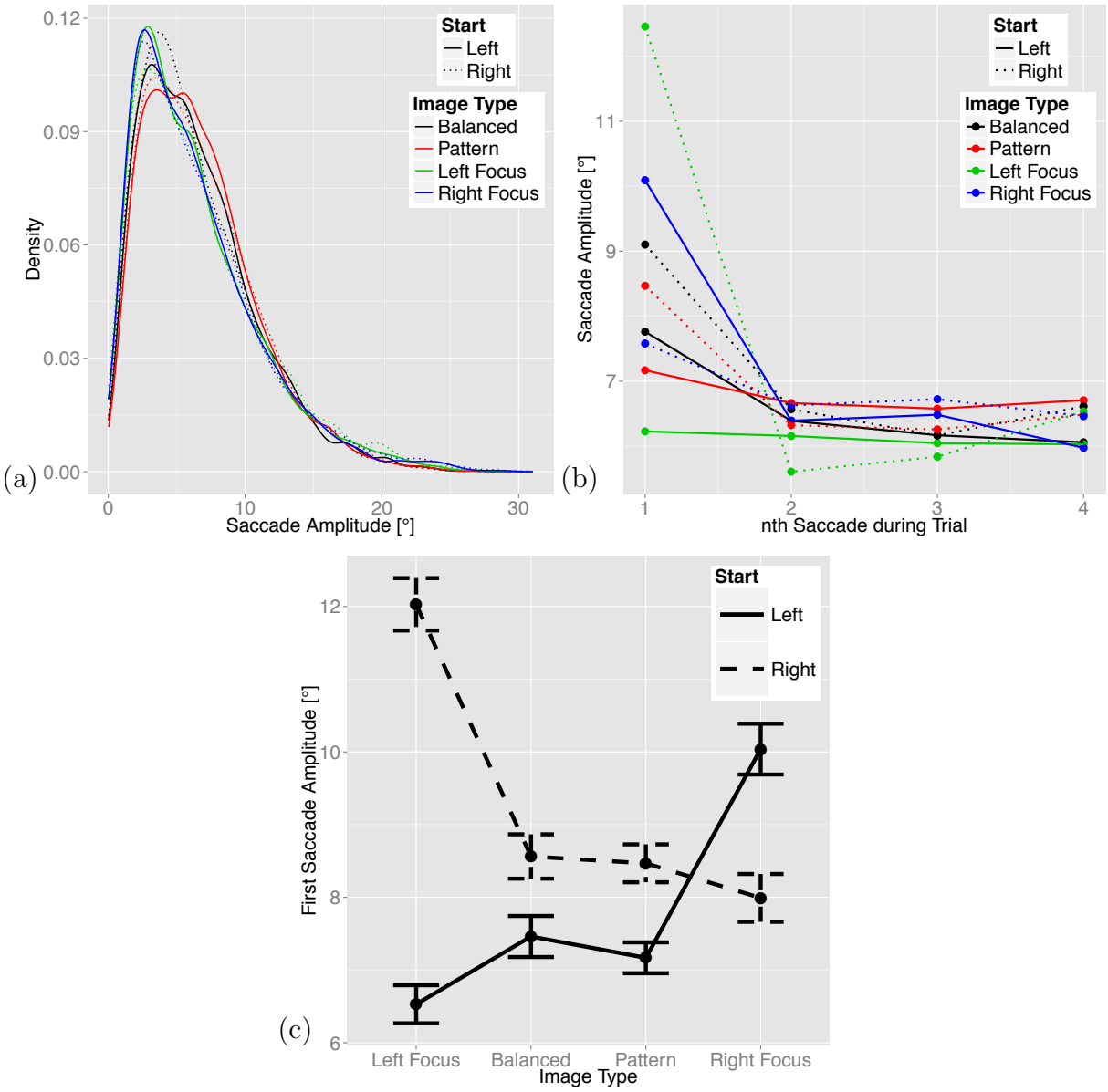

Figure 2.5: Summary statistics of saccade amplitudes. (a) Densities of all saccade amplitudes for the three images types of object-based scenes (balanced, left, and right focus) and the pattern images for left and right starting position. (b) Mean saccade amplitude for the nth saccade in each trial for all conditions. While there is a strong effect on the mean of the first saccade length, subsequent saccade amplitudes show no systematic pattern. (c) Mean values of the first saccade amplitude for the 8 different conditions. There is a strong interaction between the image type and the starting position, especially for the left focus images and the right focus images. Errorbars represent the standard error of the mean.

of the previous saccade was mapped to the origin of the coordinate system, saccade amplitude was normalized to the amplitude of the previous saccade, and the saccade orientation of the previous saccade was rotated to the right (or 180° orientation). In this representation, an endpoint at $(x; y) = (1; 0)$ corresponds to a saccade that has the same length and orientation as the previous saccade (i.e., a turning angle of 180°). The endpoint at $(x; y) = (-1; 0)$ indicates that the saccade had the same amplitude as the previous saccade , but an opposite direction, which represents a perfect return saccade (i.e., a turning angle of 0°). The high intensity at this point is consistent with earlier experiments that reported a large number of return saccades (Hooge et al., 2005; Tatler & Vincent, 2008; Smith & Henderson, 2009).

Results from our analysis of turning angles and saccade amplitudes seem - on visual inspection only - to be inconsistent with an inhibitory tagging mechanism. However, ruling out an inhibitory tagging mechanism based on these data would be premature, since inhibitory tagging could still be active, but not express in behavioral data represented in Figures 2.6a & b. Our analyses below will indicate a potential role of inhibitory tagging. Moreover, Figure 2.6c shows the same plot as Figure 2.6b, but only for the second saccade. This plot indicates return saccades appeared rarely for the second saccade, i.e. a facilitation of return back to the starting position was not observed.

2.3.3 Influence of starting position and image type on exploration behavior

The most important aim of the current study was to investigate the influence of starting position on scanpath statistics. Therefore we introduced a measure of the mean horizontal distance (MHD) to the starting position at time t, denoted by $X_{\mathrm{MHD}}(t)$ (see *Methods*). This measure was computed for each combination of image type and starting position (Fig. 2.7a). The blue horizontal line indicates the horizontal center of the image. There are three important main effects of $X_{\mathrm{MHD}}(t)$ in the plots. First, for the long term behavior in the balanced images and pattern images, $X_{\mathrm{MHD}}(t)$ approaches the midline, while there are obvious deviations for images with left or right focus.

Second, the transient behavior induced by the starting position lasts to about 3 s to 5 s (depending on condition). This observation is in strong contrast to our finding that saccade amplitudes are only affected for the first saccade, which translates into a transient phase of the mean first fixation duration, equivalent to 609.01 ms. This untypically long first fixation indicates that participants needed a long time to initiate the first saccade after disappearance of the fixation cross.

Third, after approximately 1.5 s to 2 s almost all curves cross the midline and show a local maximum of MHD. The existence of such a maximum lends support for inhibition

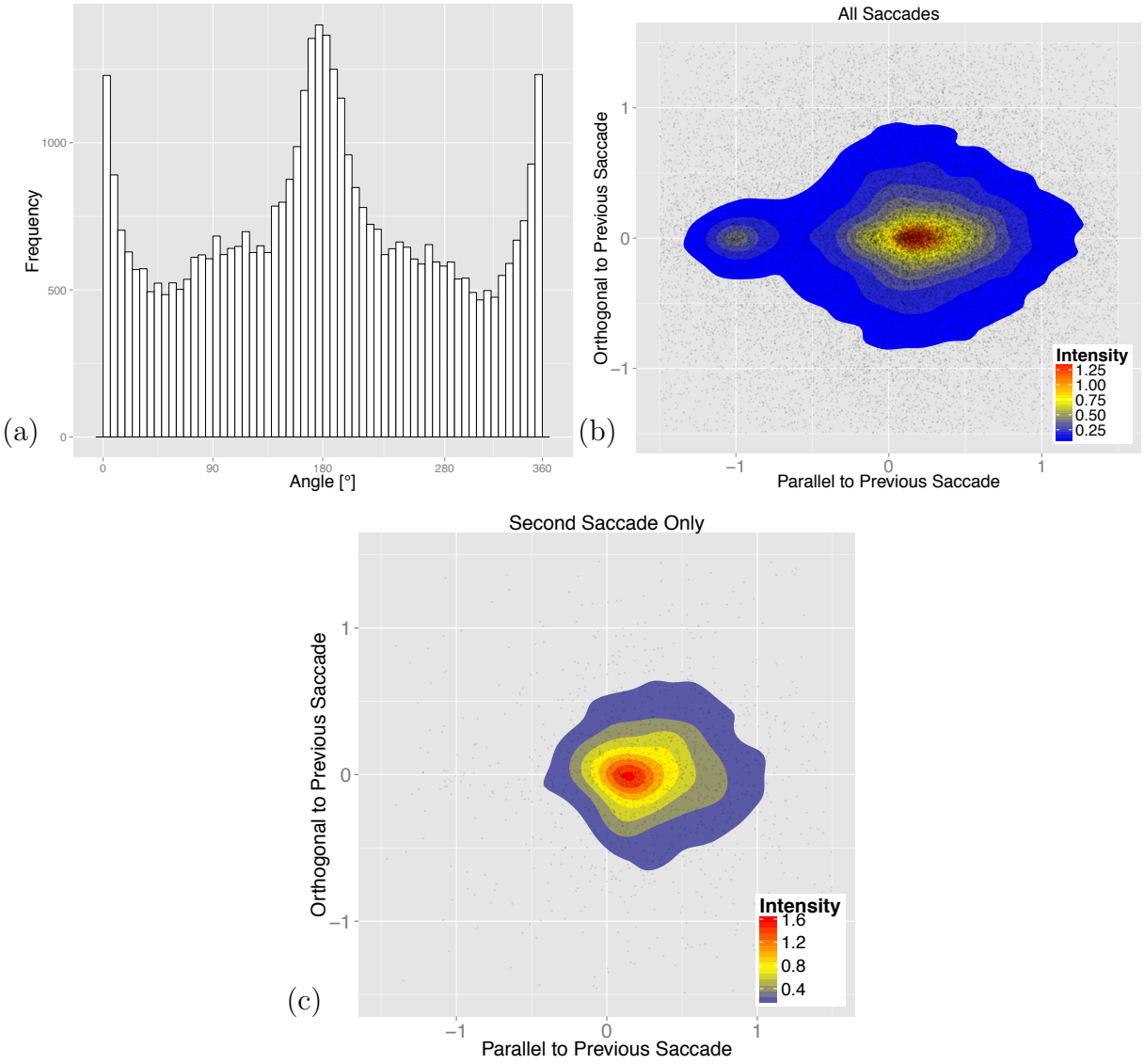

Figure 2.6: Summary statistics for saccade turning angles. (a) The distribution of angles between two successive saccades is markedly peaked at 0° (saccades that turn around) and 180° (forward saccades). (b) Plot of the relation between saccade amplitude and turning angle contingent on parameters of the previous saccade. The previous saccadic endpoints are aligned to the origin. Saccade amplitudes were normalized to one and the saccade orientations were rotated to map the endpoints of a saccade with unit length to the point $(1, 0)$. This representation shows that most saccades either travel in the same direction as the previous saccade, but with reduced saccade amplitude, or shift gaze back to the starting position of the previous saccade, i.e., the point $(-1; 0)$. (c) same as (b) but only for the first two saccades. This shows that after the long first fixation return saccades back to this position are hardly present.

at the starting position, i.e., the eye is actively driven to the opposite image side. This is most evident for the conditions in which observers started in the image side opposite to the focus (starting from the right in left-focus images and starting from the left in right-focus images), but the effect is also visible for balanced images. Additionally, an interaction between image type and starting position is visible in Figure 2.7a. When observers started in the interesting side of an image, the final MHD to the starting position is smaller than for the balanced and pattern images and for the balanced and pattern images it is smaller than if participants started in the focus image side. Graphs are cut off at $t = 6000$ ms because after approximately 5 seconds the MHD reaches an asymptotic behaviour.

Finally, we investigated the statistical reliability of our results via bootstrapping from 1 000 bootstrap samples of the 28 participants (Efron & Tibshirani, 1994). The confidence intervals (Fig. 2.7b,c) for the MHD curves $X_{\mathrm{MHD}}(t)$ were obtained by subtracting the subject mean and adding the overall mean to the samples as described by Cousineau (Cousineau, 2005; Loftus & Masson, 1994) and taking the 2.5% and 97.5% quantile of the MHD samples for the lower and upper bound. Confidence intervals show that MHD of left and right focus images differ significantly for both starting positions from the balanced images. Pattern images show almost the same MHD as balanced images. As a statistical test we computed ANOVAS to compare the mean of the balanced and pattern images (= neutral images) with the focus images. We did this at 7 different time points (0.1 s, 0.5 s, 1 s, 1.5 s, 2 s, 2.5 s and 5 s). At 0.1 s the MHD was rarely significantly ($p < 0.05$) different for the focus images and the neutral images. At 0.5 s all focus images‘ MHDs differed significantly ($p < 0.05$) from the neutral images, except for right focus images with a start on the right side. The MHD was always significantly different ($p < 0.05$) between focus and neutral images for the time points between 1 s and 2.5 s. After 5 s only some conditions show significant differences.

The difference between the images with focus (left vs. right) was significant at all time points ($p < 0.05$) after 100 ms (also at $t = 8\ s$ and $t = 10\ s$) except for inspections from the right starting position at $t = 5\ s$.

2.3.4 Comparison of experimental data with model simulations for scanpath statistics

The analysis of the time-dependence of the mean horizontal distance to the starting position uncovered at least two unexpected results, (i) the observation of long transients and (ii) an overshoot component to the image side opposite to the starting position, even in the case of balanced images. To interpret the experimental findings we calculated the same statistics for computer-generated scanpaths from two statistical models, a saccadic momentum model, a dynamical model of scene exploration as well as a combination of

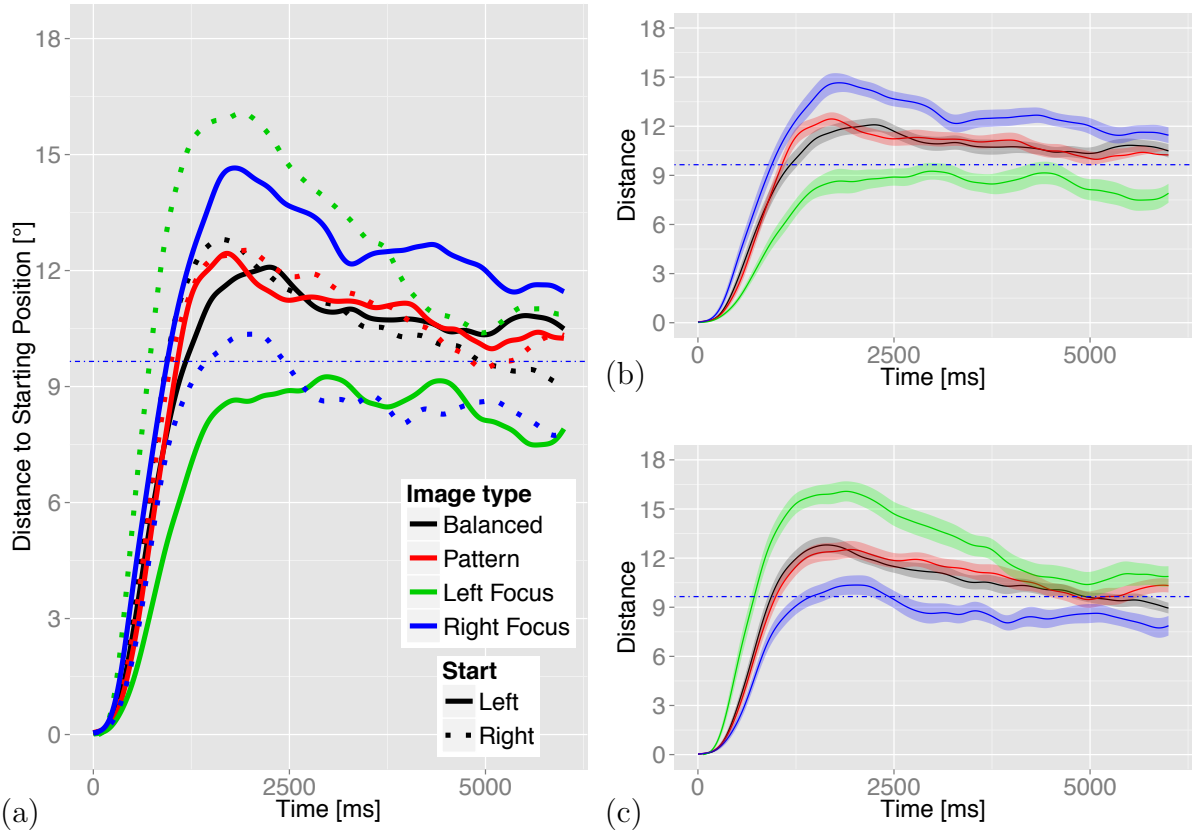

Figure 2.7: Mean horizontal distance $X_{\mathrm{MHD}}(t)$ of gaze position at time t from starting position. (a) Almost all curves show an overshoot of the mean gaze position to the image side opposite to the starting position. (b) Curves from left starting positions with bootstrap-based confidence intervals. (c) Curves from right starting positions with bootstrap-based confidence intervals.

the SceneWalk and the saccadic momentum model (see *Methods*) and compared them to the experimental data. Each model will first be described with respect to the mean horizontal distance to the starting position averaged over all experimental conditions (Fig. 2.8) and split by image type and starting position (Fig. 2.9). We will then discuss how far each model diverges from the empirically observed MHD. The time course of this deviation is plotted in Figure 2.10 and an overall comparison is presented in Figure 2.11. An example of a computer generated scanpath illustrates how each model behaves (Fig. 2.8b-g) compared to a human scanpath (Fig. 2.8a). At last we compare the distribution of angles between successive saccades of all models (Fig. 2.12) to the empirical data.

Sampling from density map

Random sampling from the density map (yellow path in Fig. 2.8) indicates that this model cannot produce an overshoot to the image side opposite to the starting position. When the starting position was oposite to the images' focus it crosses the midline (yellow path in Fig. 2.9) but stays on the opposite side afterwards. The yellow path in Figure 2.10 shows that the density model on average stays too close to the starting position until 5 seconds of observation time, because the MHD is constantly smaller than the experimental data. We only investigated the time between 1 s and 5 s because simulations of the saccadic momentum model equal the empirical data until 999 ms and after 5 s the empirical MHD curve reaches a stable value. We pooled the eight conditions from Figure 2.9 into the three conditions neutral images, start within the focus side and start opposite to the focus. There were no systematic differences between different conditions within these groups. The simluated scanpath of the density model (Fig. 2.8b) coveres similar locations as human observers but fails to produce the systematic scanning behaviour. This claim is supported by Figure 2.12. The angle distribution of the density model does not resemble human behaviour. The density model generated almost no forward saccades (i.e. 180°) but a very large amount of backward saccades. This agrees with the findings that a memoryless system produces more return saccades than present in the data (Bays & Husain, 2012).

Gaussian Model

Though the Gaussian-weighted model is psychologically more plausible than the density model because of its limited attentional span, it performs even worse with respect to the MHD (light blue path in Fig. 2.8 & 2.9). It leaves the starting position even slower than the density model due to the limited attention span. The deviation of the MHD from the experimental data (Fig. 2.10) is always negative. The scanpath of the gauss model (Figure 2.8c) indicates that the limited attention span and the absence of inhibition

leads to rather small saccades. Thus the scanpath of the gaussian model often covers a smaller area of the image, than human observers. The angular distribution shows too little forward saccades (Fig. 2.12). The peak of return saccades is less pronounced than for the density model but is still evident. This is due to the fact, that the model cannot cross image boundaries and thus, when a fixation is placed close to the border, cannot produce forward saccades.

Saccadic Momentum Model

Although the saccadic momentum model contains the same initial two fixations as the data, it cannot reproduce the overshoot from the data (dark blue path in Fig. 2.8 & 2.9). The MHD is almost constantly lower than that of the human data (Figure 2.10) and it's MHD is even further away from the data in two of the three conditions than that of the density model (Fig. 2.11). Figure 2.8d shows a scanpath that is similar to human data with respect to saccade lengths and angles but does neither capture the locations looked at nor dynamical aspects of human data. The angular distribution resembles that of the experimental data (Fig. 2.12). Thus, the model reproduces the distribution of angles between successive saccades but fails to produce a similar MHD as the data, because it stays too long on the image side of the initial fixation.

SceneWalk model

In contrast to the other models, the dynamical SceneWalk model (Engbert et al., 2015) reproduces the overshoot component of the MHD curves in the time interval between 1.5 s and 2 s (pink path in Fig. 2.8 & 2.9). The SceneWalk model uses inhibitory tagging that drives the eyes away from the starting position by suppressing the selection of saccade targets close to the initial fixation position. The MHD produced by the SceneWalk model is closer to the empirical MHD than models without inhibitory tagging (Fig. 2.10) when the starting position was not in the side of the images' focus. If starting positions were on the focus side, the overshoot produced by the SceneWalk model was too strong (Fig. 2.9).

The scanpath produced by the SceneWalk model (Figure 2.8e) resembles a typical human scanpath, because it doesn't stick to any locations and inspects important image parts more thoroughly. The angle distribution of the SceneWalk model however does not resemble human data (Figure 2.12). It shows a similar distribution as the density and the gaussian model but the peak of the return saccades is less pronounced.

Adjusted SceneWalk model

Since model parameters of the SceneWalk model were taken from the published version and not adjusted to the current experimental data, we changed the exponent of the inhibition map from $\gamma = .3$ to .2 (see Eq. 2.2.3 in Methods) in a second simulation (dark red path in Fig. 2.8 & 2.9). This parameter controls the inhibition map and influences the amount of aggregation (or clustering) of realized gaze positions. We see that the overshoot of the MHD curve of the SceneWalk model with an adjusted exponent of the inhibition map is in good agreement with the overshoot observed in the experimental data (Fig. 2.10). These simulations suggest that the overshoot produced by the model is primarily caused by the inhibitory tagging mechanism. The angle distribution of the adjusted SceneWalk model (Fig. 2.12) as well as the computed scanpath (Fig. 2.8f) showed similar characteristics as the orignial SceneWalk model. Because this model was adjusted post hoc, we did not statistically compare it to the other models.

SceneWalk Model + Saccadic Momentum

With an additional saccadic momentum mechanism, the SceneWalk model still produces the overshoot seen in the data (red path in Fig. 2.8 & 2.9). As in the original SceneWalk model, the overshoot in MHD of the augmented model is sometimes too strong (Fig. 2.10), especially when the starting position is within the focus side of the image. The scanpath produced by this model looks similar to the human scanpath (Fig. 2.8g). The overshoot of MHD is reproduced by the model and the angle distribution is very similar to the human data (Fig. 2.12).

Statistical model comparison

We computed an ANOVA to statistically compare the performance of the SceneWalk model to other models. We compared the mean deviation of MHD between models and experimental data in the interval from 1 s to 5 s (Figure 2.11). Because the adjusted SceneWalk model was hand tuned post hoc, we will only statistically compare the original SceneWalk to the other models. In neutral images the SceneWalk model performs significantly better than the density model ($F(1, 54) = 18.77, p < 0.001$), the gaussian model ($F(1, 54) = 91.49, p < 0.001$) and the saccadic momentum model ($F(1, 54) = 23.77, p < 0.001$). If the initial fixation position was on the side of the scene focus the SceneWalk model did not differ significantly from the density model ($F(1, 54) = 1.398, p = 0.242$) or the saccadic momentum model ($F(1, 54) = 1.736, p = 0.193$) but performed significantlyy better than the gaussian model ($F(1, 54) = 7.13, p < 0.01$). For a starting position opposite to the focus the MHD produced by the SceneWalk model differed significantly less from the empirical MHD than the density model ($F(1, 54) = 17.63, p < 0.001$),

the gaussian model ($F(1,54) = 116, p < 0.001$) and the saccadic momentum model ($F(1,54) = 22.55, p < 0.001$). Comparing the original SceneWalk model to the SceneWalk + saccadic momentum model did not show significant differences ($p < 0.05$) for any condition.

2.4 Discussion

In an eye tracking experiment we investigated the influence of experimentally manipulated starting positions on scanpath behavior in human observers. The most important effects were observed in the temporal evolution of the mean horizontal distance (MHD) to the starting position. First, we found unexpectedly long transients in mean eye position. It took up to 5 seconds for gaze of human observers to reach the final average fixation position. This is a lot longer than the saccade amplitude effects, which were limited to the very first saccade of an observers' scanpath. Second, for almost all experimental conditions the MHD over time is characterized by a strong overshoot of the midline into the image side opposite to the starting position before reaching a stable value. This effect lends support to a foraging strategy that actively moves the gaze to unexplored image regions although on a shorter time scale, a high number of return saccades suggests the opposite.

Next, we analyzed computational models that incorporate mechanisms of eye movement control to produce human scanpaths. Random sampling from the empirical fixation map (i.e., assuming a 'perfect' fixation density model) does not replicate human behavior, since the overshoot to the opposite side of the image cannot be reproduced and the distribution of angles between successive saccades did not resemble human behaviour. Considering that this density model is a 'perfect' fixation density model this is quite remarkable, because it shows that even if we can perfectly predict fixation locations, human eye movement behaviour is not reproduced by default. Additionally, such a model is psychologically highly implausible because of the missing effect of degraded visual processing towards the periphery of the visual field. However, an augmented model, i.e., a combination of the density map with a gaussian attention window representing the fall-off of visual processing to the periphery, performs even worse compared to random sampling from the empirical map. We conclude from these results that an active mechanism driving the eyes away from the starting position is necessary to explain scanpath statistics as the time-dependence of mean horizontal distance.

Given the above experimental results, we were looking for potential principles of eye guidance that drive the trajectory faster away from the current fixation position than a simple random process. We investigated two principals in computational models: saccadic momentum (Smith & Henderson, 2009; Wilming et al., 2013) and spatial inhibitory

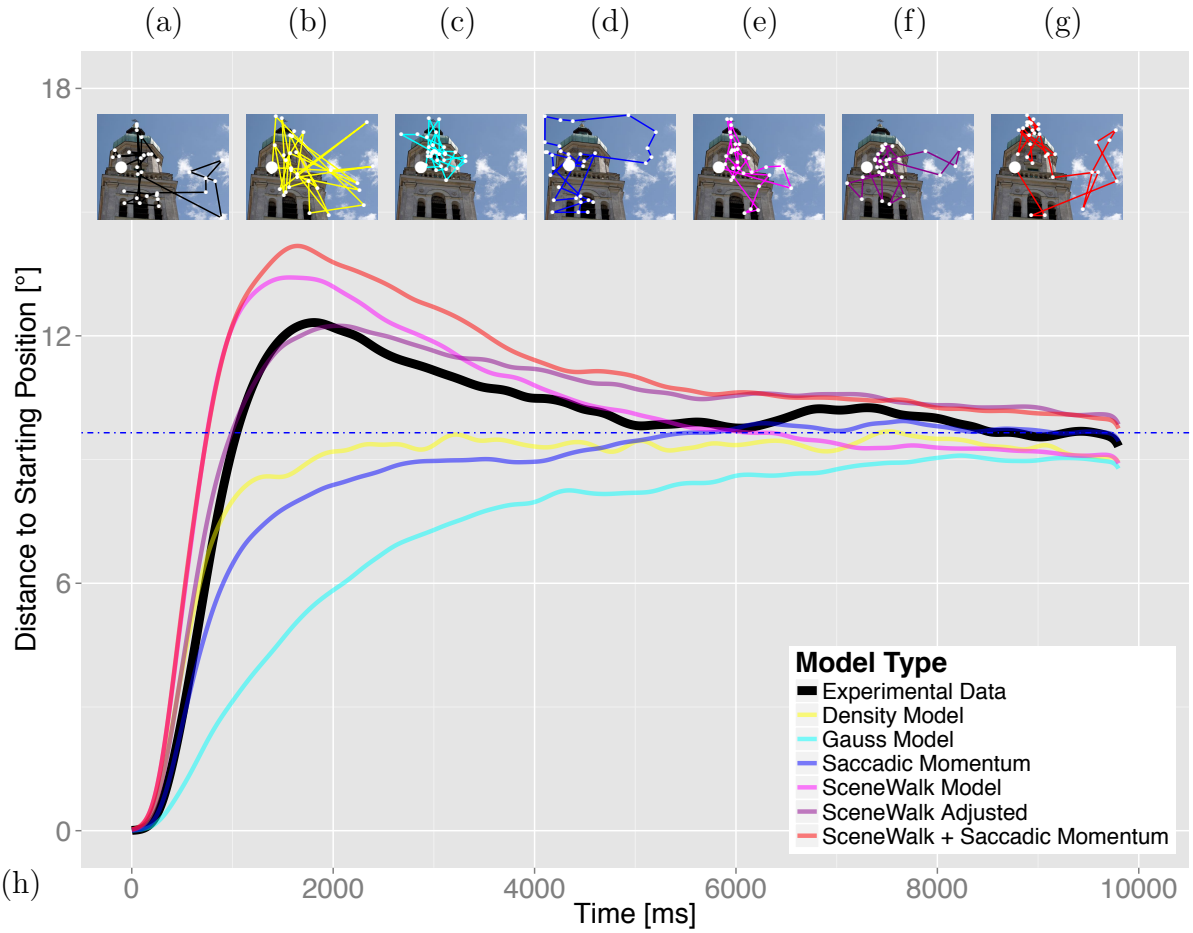

Figure 2.8: Comparison of mean horizontal distance of gaze from starting position for the experimental data and the computational models. Examples of scanpaths for (a) experimental data, (b) random sampling from density map, (c) gaussian weighted random sampling from density map, (d) saccadic momentum model, (e) SceneWalk model (Engbert et al., 2015) based on target selection from dynamic activation maps, (f) SceneWalk model with an adjusted exponent of the inhibition map and (g) SceneWalk + Saccadic Momentum Model. (h) Mean horizontal distance $X_{\mathrm{MHD}}(t)$ of gaze position at time t shows that the qualitative behavior in the experimental data with an overshoot component to the image side opposite to the starting position is reproduced by the SceneWalk models that use inhibitory tagging as a driving mechanism.

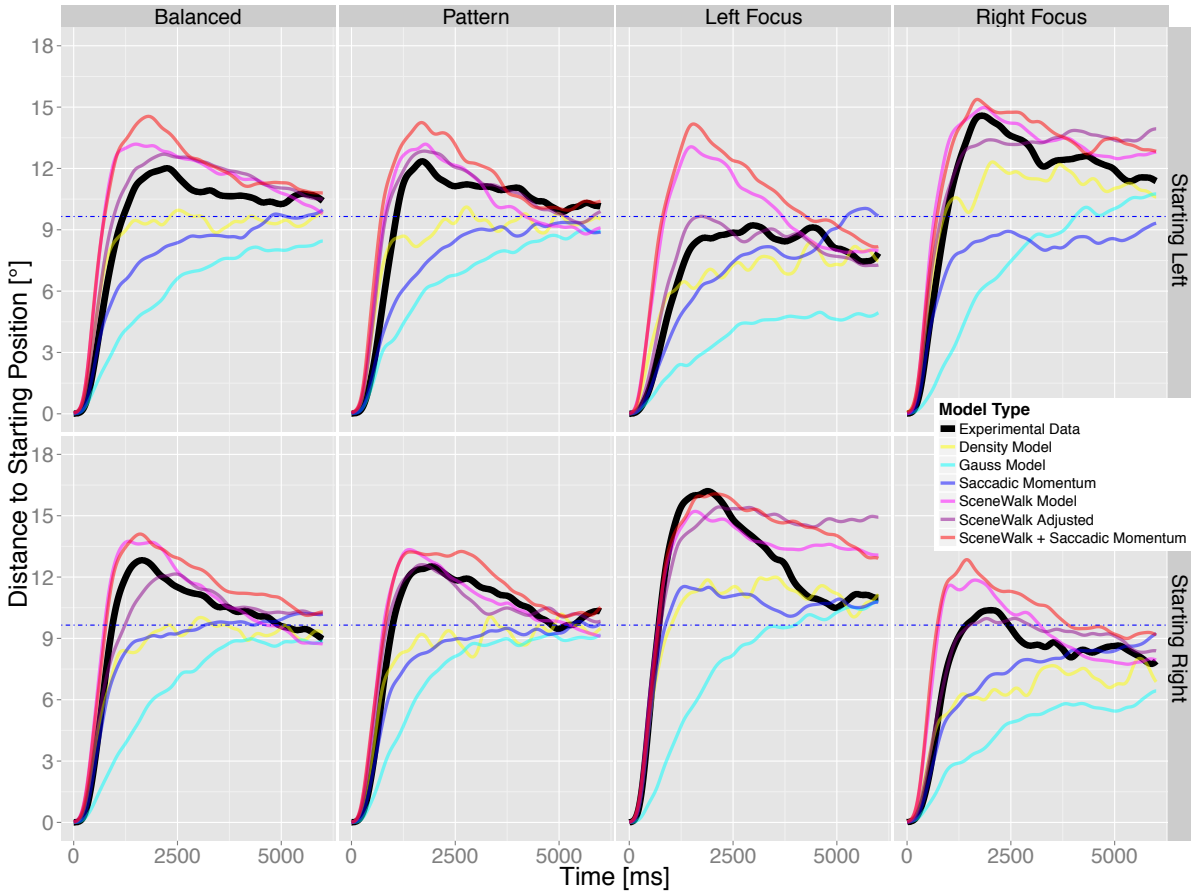

Figure 2.9: Mean horizontal distance to the starting position for all 8 combinations of image type and starting position, for the 4 different scanpath models and experimental data. In all but one condition (left-focus image with left starting position), an overshoot of the mean position to the image side opposite to the starting position is visible in the experimental data. This overshoot was reproduced by the dynamical SceneWalk models that implement inhibitory tagging.

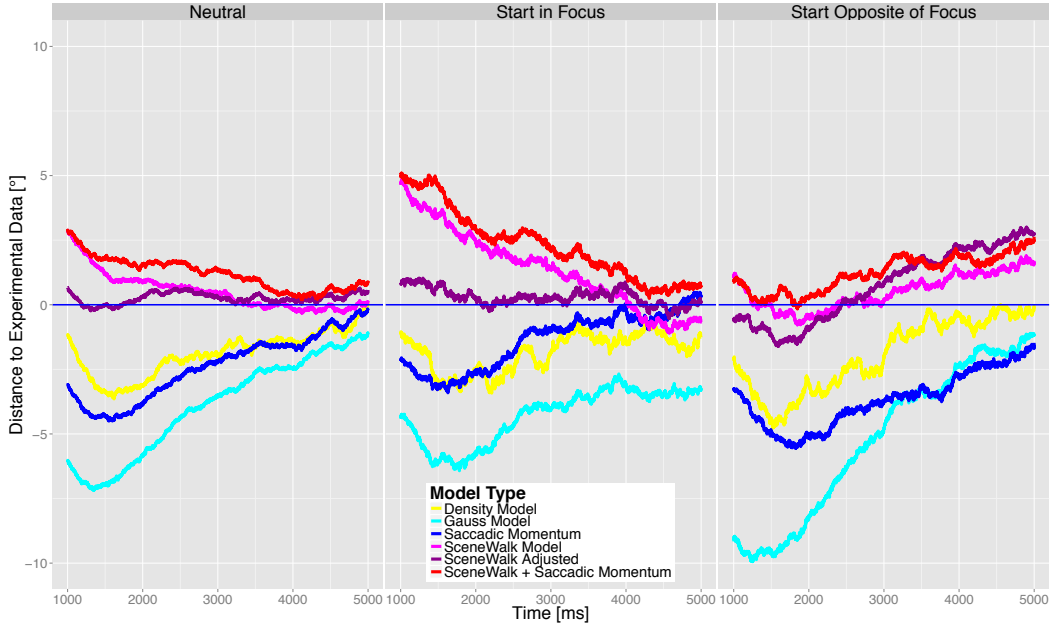

Figure 2.10: Difference between the MHD produced by the models and the experimental data between 1 and 5 seconds. The SceneWalk models are often above the zero-error line, indicating that the overshoot of the MHD is too strong. The other models are always below this line, indicating that they leave the starting position slower than human observers.

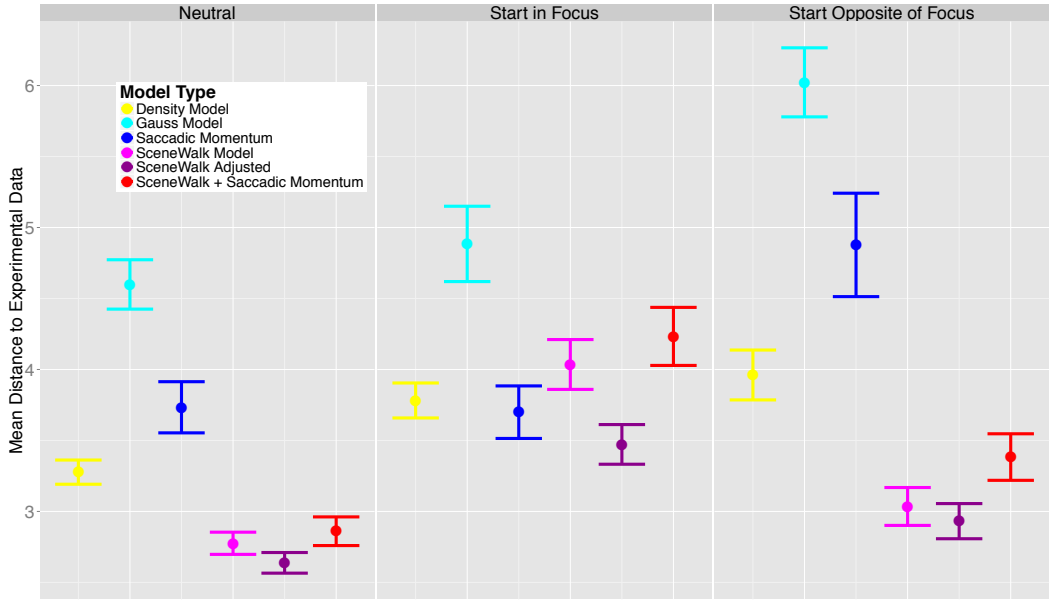

Figure 2.11: Mean absolute difference between the MHD produced by the models and the experimental data between 1 and 5 seconds. For neutral images and if the starting position was opposite of the focus the SceneWalk models perform better than the other models. If the initial fixation position lies in the focus side of an image there is no significant difference between the orignial SceneWalk model and the other predictive models (density model, gaussian model, saccadic momentum model).

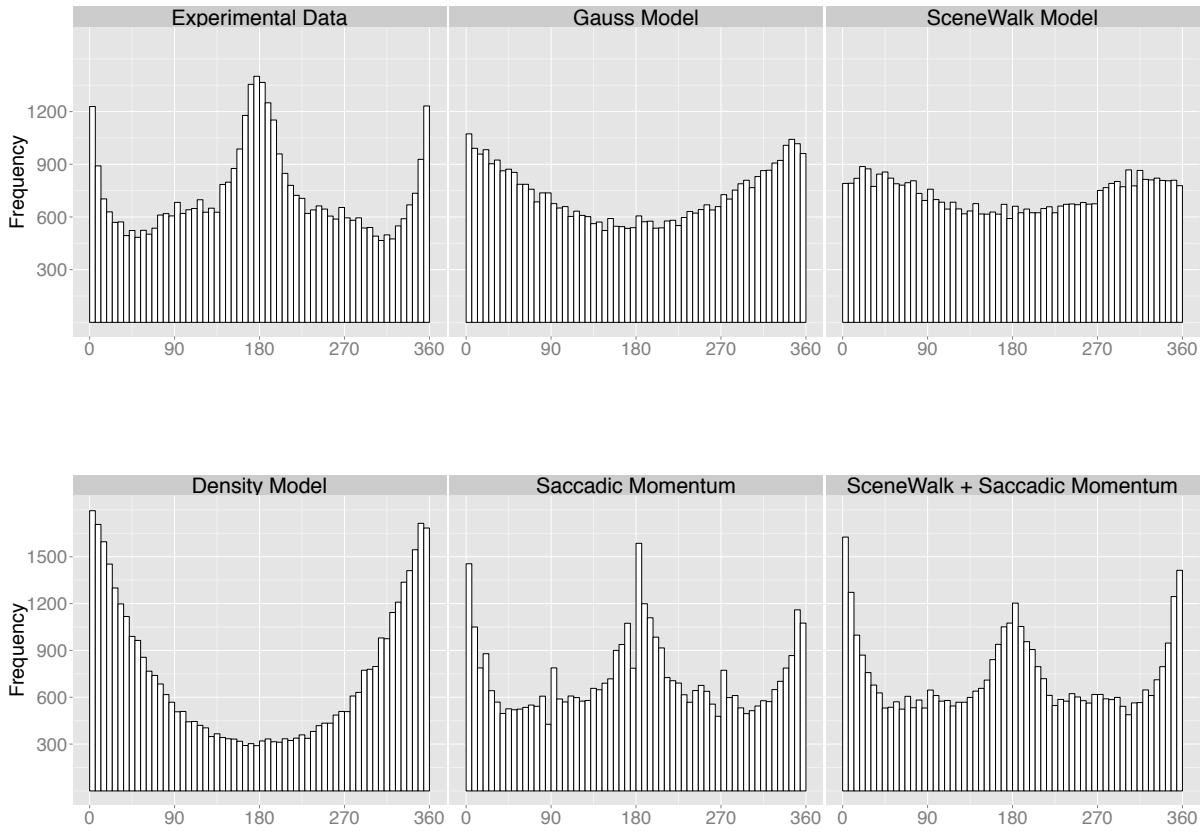

Figure 2.12: Distribution of angles between successive saccades for all models (see Fig. 2.6a). The density model, the gauss model and the original and adjusted SceneWalk model can not reproduce the angle distribution of the experimental data. A saccadic momentum model that is based on the angle distribution of the data can reproduce it as well as the SceneWalk model augmented with a saccadic momentum mechanism.

tagging (Itti et al., 1998; Le Meur & Liu, 2015).

We designed a model with saccadic momentum that samples from the joint probability distribution of saccade angles and amplitudes and keeps the first saccade as observed in the experiments. As a trivial result, the first two fixations in each trial from the simulations fit the experimental data better than any other model. However, the model did not reproduce the overshoot component to the opposite image side.

We also used the SceneWalk model (Engbert et al., 2015), a dynamical model for eye-movement control in scenes that reproduces first- and second-order statistics, i.e., densities of fixation locations and clustering of fixation locations, respectively. The SceneWalk model uses inhibitory tagging, a mechanism motivated by the findings on inhibition of return (Posner & Cohen, 1984; Posner et al., 1985; Klein, 1988). We demonstrated that the SceneWalk model generates the overshoot effect for MHD via inhibitory tagging in contrast to the two random-sampling models or the saccadic momentum model. With the parameters that were fitted from a different experiment, the SceneWalk model produced MHD curves that were more similar to the curves computed from the experimental data than all other models. Reducing the exponent of the inhibition map alternated the overshoot, in particular when participants started in the focus side of an image. To account for the distribution of angles between successive saccades we added a map that weights possible future fixation locations with respect to the probability of angles between successive saccades. This model reproduced the observed angle distribution (Fig. 2.12) and produced the overshoot that was not observed in a simple saccadic momentum model (see Fig. 2.8). Although the empirical angle distribution shows a peak at return saccades, a facilitation of return (Smith & Henderson, 2009) back to the starting position was not observed (Fig. 2.6c).

We investigated models using an inhibitory tagging mechanism and a saccadic momentum mechanism. The overshoot to the image side that is observed in human data is reproduced by all models that implement inhibitory tagging. The angle distribution between two succesive saccades is produced by all models that use a saccadic momentum mechanism. Without any of these mechanism, both measures fail to be reproduced by a model. This shows that the distribution of angles with a large amount of return saccades and an inhibitory tagging mechanism are not necessarily a contradiction but reproduce certain spatial and temporal dependencies of human scanpaths when applied together in a model.

Inspections of a left-focus image from a starting position on the left show different dynamics of the mean horizontal distance compared to all other conditions. This could be due to a stronger directional bias in left-focus images than in right-focus images in our experimental material. It is also possible that there is a general tendency to first look at the left image side and then scan to the right—a tendency that has been found

earlier in scene viewing (Dickinson & Intraub, 2009; Ossandón et al., 2014), pattern exploration (Abed, 1991) and face viewing (Guo, Meints, Hall, Hall, & Mills, 2009)—which is congruent to the reading direction of our participants. A dynamical model of eye guidance might perform better with an additional Bayesian-type prior probability implementing a leftward bias and a center bias for initial saccades. Thus, our results emphasize the need for more advanced dynamical models of scanpath generation.

2.5 Conclusion

The experimental manipulation of starting position exerts a strong and long lasting influence on scanpaths during scene exploration. Using computational models, we demonstrate that a model with inhibitory tagging can explain the mean overshoot of gaze position to the image side opposite to the starting position whilst simple statistical models as well as a saccadic momentum model without inhibitory tagging do not reproduce this overshoot. In addition, even if we are able to predict a perfect fixation density model, we are still far from predicting spatial and temporal dependencies between successive fixations. Our results lend support to inhibitory tagging as a dynamical principle of saccade planning during scene viewing.

2.6 Acknowledgements

This work was supported by Deutsche Forschungsgemeinschaft (grants EN 471/13–1 and WI 2103/4–1 to R. E. and F. A. W., resp.).

2.7 Appendix

Because the experimental design turned out to be difficult for our participants, many trials had to be repeated. Figures 2.13-2.17 represent Figures 2.5–2.9 if only trials without a repeated fixation check are taken into account. All results are very similar and no systematic difference was observed between trials with and without a repeated fixation test.

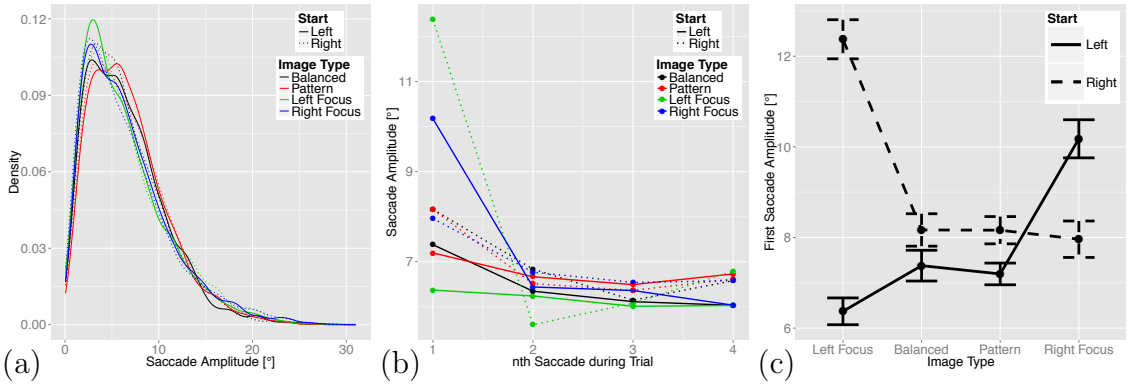

Figure 2.13: Summary statistics of saccade amplitudes for trials without repetition of the fixation test. (a) Densities of all saccade amplitudes for the three images types of object-based scenes (balanced, left, and right focus) and the pattern images for left and right starting position. (b) Mean saccade amplitude for the nth saccade in each trial for all conditions. While there is a strong effect on the mean of the first saccade length, subsequent saccade amplitudes show no systematic pattern. (c) Mean values of the first saccade amplitude for the 8 different conditions. There is a strong interaction between the image type and the starting position especially for the left focus images and the right focus images. Errorbars represent the standard error of the mean.

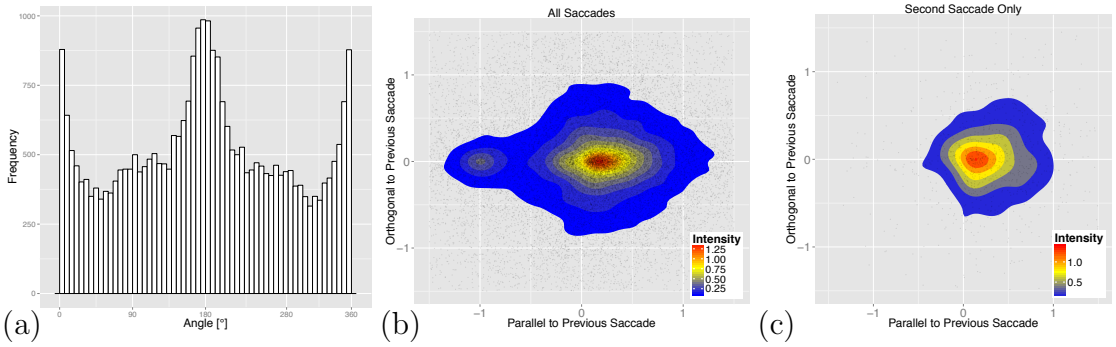

Figure 2.14: Summary statistics for saccade turning angles for trials trials without repetition of the fixation test. (a) The distribution of angles between two successive saccades is markedly peaked at 0° (saccades that turn around) and 180° (forward saccades). (b) Plot of the relation between saccade amplitude and turning angle contingent on parameters of the previous saccade. The previous saccadic endpoints are aligned to the origin. Saccade amplitudes were normalized to one and the saccade orientations were rotated to map the endpoints of a saccade with unit length to the point $(1, 0)$. This representation shows that most saccades either travel in the same direction as the previous saccade, but with reduced saccade amplitude, or shift gaze back to the starting position of the previous saccade, i.e., the point $(-1; 0)$. (c) same as (b) but only for the first two saccades. This shows that after the long first fixation return saccades back to this position are hardly present.

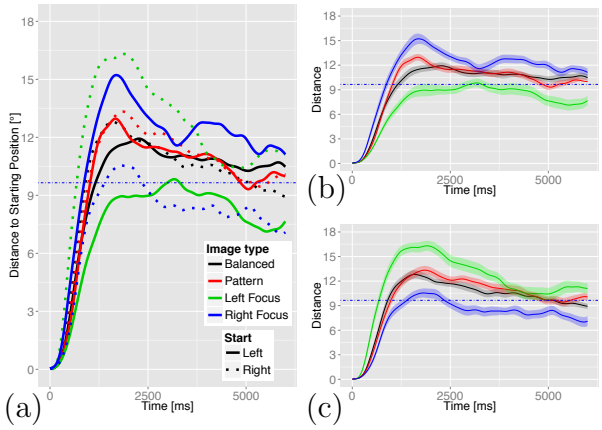

Figure 2.15: Mean horizontal distance $X_{\mathrm{MHD}}(t)$ of gaze position at time t from starting position for trials without repetition of the fixation test. (a) Almost all curves show an overshoot of the mean gaze position to the image side opposite to the starting position. (b) Curves from left starting positions with bootstrap-based confidence intervals. (c) Curves from right starting positions with bootstrap-based confidence intervals.

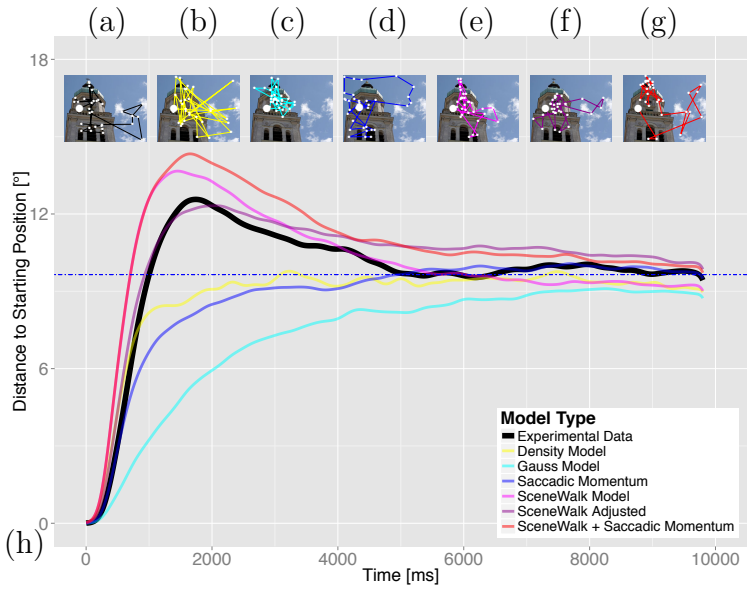

Figure 2.16: Comparison of mean horizontal distance of gaze from starting position for the experimental data and the computational models for trials without repetition of the fixation test. Examples of scanpaths for (a) experimental data, (b) random sampling from density map, (c) gaussian weighted random sampling from density map, (d) saccadic momentum model, (e) SceneWalk model (Engbert et al., 2015) based on target selection from dynamic activation maps, (f) SceneWalk model with an adjusted exponent of the inhibition map and (g) SceneWalk + Saccadic Momentum Model. (h) Mean horizontal distance $X_{\mathrm{MHD}}(t)$ of gaze position at time t shows that the qualitative behavior in the experimental data with an overshoot component to the image side opposite to the starting position is reproduced by the SceneWalk models that use inhibitory tagging as a driving mechanism.

Figure 2.17: Mean horizontal distance to the starting position for all 8 combinations of image type and starting position, for the scanpath models and experimental data for trials without repetition of the fixation test. In all but one condition (left-focus image with left starting position), an overshoot of the mean position to the image side opposite to the starting position is visible in the experimental data. This overshoot was reproduced by the dynamical SceneWalk models that implement inhibitory tagging.

Chapter 3

The temporal evolution of the central fixation bias in scene viewing

Lars O. M. Rothkegel[1*], Hans A. Trukenbrod[1], Heiko H. Schütt[1,2], Felix A. Wichmann[2–4], and Ralf Engbert[1]

[1]University of Potsdam, Germany
[2]Eberhard Karls University Tübingen, Germany
[3]Bernstein Center for Computational Neuroscience Tübingen, Germany
[4]Max Planck Institute for Intelligent Systems, Tübingen, Germany

Running head: Central fixation bias

Published 2017 in *Journal of Vision, 17* (13) 3:1–18.
doi: 10.1167/17.13.3

Abstract

When watching the image of a natural scene on a computer screen, observers initially move their eyes toward the center of the image — a reliable experimental finding termed *central fixation bias*. This systematic tendency in eye guidance likely masks attentional selection driven by image properties and top-down cognitive processes. Here, we show that the central fixation bias can be reduced by delaying the initial saccade relative to image onset. In four scene-viewing experiments we manipulated observers' initial gaze position and delayed their first saccade by a specific time interval relative to the onset of an image. We analyzed the distance to image center over time and show that the central fixation bias of initial fixations was significantly reduced after delayed saccade onsets. We additionally show that selection of the initial saccade target strongly depended on the first saccade latency. A previously published model of saccade generation was extended with a central activation map on the initial fixation whose influence declined with increasing saccade latency. This extension was sufficient to replicate the central fixation bias from our experiments. Our results suggest that the central fixation bias is generated by default activation as a response to the sudden image onset and that this default activation pattern decreases over time. Thus, it may often be preferable to use a modified version of the scene viewing paradigm that decouples image onset from the start signal for scene exploration to explicitly reduce the central fixation bias.

3.1 Introduction

How humans visually explore natural scenes depends on multiple factors. Eye movements are influenced by low level image properties (e.g., chromaticity, orientation, luminance, and color contrast; Itti et al., 1998; Torralba, 2003; Le Meur, Le Callet, Barba, & Thoreau, 2006) as well as higher level cognitive processes like the observers' scene understanding (Loftus & Mackworth, 1978; Henderson et al., 1999), task (Yarbus et al., 1967; Castelhano & Henderson, 2008a), or probability of reward (Hayhoe & Ballard, 2005; Tatler et al., 2011). Besides low-level image features and high-level cognition, *systematic tendencies* have a strong impact on how humans look at pictures (Tatler & Vincent, 2009; Le Meur & Liu, 2015). A dominant systematic tendency in natural scene viewing is the *central fixation bias* (CFB; Buswell, 1935; Tatler, 2007; Tseng, Carmi, Cameron, Munoz, & Itti, 2009). Regardless of stimulus material (Tatler, 2007; Tseng et al., 2009), head position (Vitu, Kapoula, Lancelin, & Lavigne, 2004), initial fixation position (Tatler, 2007; Bindemann, Scheepers, Ferguson, & Burton, 2010), or image position (Bindemann, 2010), the eyes tend to initially fixate close to the center of an image when presented to a human observer on a computer screen. After several explanations of the CFB had been ruled out, two hypotheses remained.

First, the image center might be the best location to maximize information extraction from scenes (Najemnik & Geisler, 2005; Tatler, 2007) – at least for typical photographs found in image databases and on the internet (c.f; Wichmann, Drewes, Rosas, & Gegenfurtner, 2010). Second, the center provides a strategic advantage to start the exploration of an image (Tatler, 2007). Because real-world visual input does not suddenly appear and peripheral information of an upcoming stimulus is usually available, the CFB might be a laboratory artifact to some degree. Also, natural visual stimuli do not have rigid boundaries like a computer screen. A reduction of the CFB in mobile eye tracking data ('t Hart et al., 2009; Ioannidou, Hermens, & Hodgson, 2016) supports this idea.

A previous study from our lab resulted in a strong reduction of the CFB on initial fixations compared with similar experiments. In this study we manipulated the initial fixation by requiring participants to maintain fixation on a starting position close to the border of the screen for 1 s (Rothkegel, Trukenbrod, Schütt, Wichmann, & Engbert, 2016). In addition, some images in this study had asymmetric conspicuity distributions, with interesting or *salient* image parts on either side of the image, but less so in the center. Thus, the reduction of the CFB in our scene-viewing experiment could have been generated by three aspects: extreme initial starting positions, delayed initial saccades, and the saliency bias of the images we used.

To investigate the principles underlying the reduced CFB, we designed and analyzed four experiments, in which observers started exploration from different positions within an

image and were required to maintain fixation for various time intervals after image onset (pretrial fixation time). Our study used the images investigated in the most frequently cited paper on the central fixation bias (Tatler, 2007), to exclude any influence of the images on the reduction of the CFB.

We hypothesized that (a) a forced prolonged initial fixation decouples image onset from the signal to start exploration and leads to a reduced CFB on the second fixation which in turn reduces the bias on subsequent fixations (due to the short saccade amplitudes of humans during scene perception; Tatler & Vincent, 2008) and that (b) the magnitude of the reduction varies with the duration of the prolonged initial fixation.

Here, we show that the CFB of early eye movements can be reduced by dissociating initial eye movements from a sudden image onset by 75 ms and more. Increasing the delay of the initial response by more than 250 ms produced only marginal differences. In addition, we show that the initial saccade latency predicts the strength of the CFB on a trial-by-trial basis. The pretrial fixation time primarily assures that the initial fixation is long enough to avoid a strong orienting response to the center of an image. By implementing these results in a previously published model of saccade generation (Engbert et al., 2015) we were able to reproduce the influence of saccade latency on the CFB as well as the qualitative progression of the CFB over time.

3.2 General methods

3.2.1 Stimuli

A set of 120 images was presented on a 20-in. CRT monitor (Mitsubishi Diamond Pro 2070: frame rate 120 Hz, resolution 1,280 × 1,024 pixels; Mitsubishi Electric Corporation, Tokyo, Japan) in Experiments 1, 2 and 4 and on a different 20-in. CRT monitor in Experiment 3 (Iiyama Vision Master Pro 514: frame rate 100 Hz, resolution 1,280 × 1,024 pixels; Iiyama, Nagano, Japan). The images were the same as in Tatler's (2007) original study on the central fixation bias. Images were indoor scenes (40 images), outdoor scenes with manmade structures present (e.g., urban scenes; 40 images), and outdoor scenes with no manmade structures present (40 images). Images were taken using a Nikon D2 digital SLR using its highest resolution (4 megapixel). All pictures had a size of 1,600×1,200 pixels. For the presentation during the experiment, images were converted to a size of 1,200 × 900 pixels and centered on a screen with gray borders extending 64 pixels to the top/bottom and 40 pixels to the left/ right of the image. In Experiments 1, 2, and 4 the images covered 31.1° of visual angle in the horizontal and 23.3° in the vertical dimension. In Experiment 3 images covered a larger proportion of the visual field with 36.25° of visual angle in the horizontal and 27.20° in the vertical dimension due to a

reduced viewing distance.

3.2.2 Participants

Participants were students of the University of Potsdam and of nearby high schools. Number of participants will be reported for each experiment separately. They received credit points or a monetary compensation of 8 Euro for their participation in any of the four experiments. The average duration of one experimental session was 40-45 min. All participants had normal or corrected-to-normal vision. The work was carried out in accordance with the Declaration of Helsinki. Informed consent was obtained for experimentation by all participants.

3.2.3 General procedure

Participants were instructed to position their heads on a chin rest in front of a computer screen at a viewing distance of 70 cm (60 cm in Exp. 3). Eye movements were recorded binocularly (monocularly in Experiment 3) using an EyeLink 1000 video-based-eye tracker (desktop mount system for Experiments 1,2, and 4 and tower mount system for Exp. 3; SR Research, Osgoode, ON, Canada) with a sampling rate of 500 Hz (1000 Hz in Exp. 3 and downsampled to 500 Hz for our analysis). Trials began with a black fixation cross presented on a gray background. After successful fixation, an image was presented. After onset of the image, the fixation cross remained visible on top of the image for a variable duration. We refer to this duration as the pretrial fixation time. Participants were instructed to keep their eyes on the fixation cross until it disappeared. If participants moved their eyes before the pretrial fixation time elapsed, a mask of Gaussian white noise was displayed and the trial started anew with the initial fixation check. After successful initial fixation, participants were instructed to explore the scene freely for 5 s in all experiments. Experiments were run with the MATLAB software (MATLAB, 2015) using the Psychophysics (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007) and EyeLink (F. W. Cornelissen, Peters, & Palmer, 2002) toolboxes.

3.2.4 Data analysis

Data preprocessing and saccade detection

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

3.2.5 Distance to center over time

We computed the mean distance of the eye position to the image center DTC as a function of pretrial fixation time (T). This was computed as follows

$$DTC_T = \frac{1}{m \cdot n} \sum_{j=1}^{n} \sum_{k=1}^{m} ||x_{jk}(t) - x_{center}|| \,, \tag{3.1}$$

where $x_{jk}(t)$ indicates gaze position of participant j on image k at time t and x_{center} indicates the image center. The vertical bars indicate the Euclidian distance from the center for each gaze position. As a continuous-time measure, we computed the DTC of each sample of the eye position time series. In this representation, a larger DTC indicates a less pronounced CFB and vice versa. For all experiments we visualized the mean $DTC(t)$ to the image center for the entire 5-s observation window for each pretrial fixation time. The observation window started at $t = 0$ with the disappearance of the fixation marker. All figures were created with the ggplot2 package (Wickham, 2009) of the R-Language of Statistical Computing (R Core Team, 2014).

Influence of the initial fixation on the second fixation

The pretrial fixation time influenced the DTC on early fixations. To further investigate this influence, we plot the DTC of the second fixation as a function of overall saccade latency from image onset. We computed linear mixed models (Bates, Mächler, Bolker, & Walker, 2015) with initial saccade latency and pretrial fixation time as fixed effects, the DTC of the second fixation as the dependent variable and an intercept for participants and images as random factors. To compute the models, we transformed DTC with the boxcox function of the R package MASS (Venables & Ripley, 2002) to follow a normal distribution. We obtained significance levels with the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2013). Contrasts were defined as sum contrasts. This means that each pretrial fixation time is compared with the overall mean of distance to center. To be able to compare the different factor levels with the overall mean, the highest pretrial fixation time in each experiment was left out. In all experiments we excluded saccades with latencies smaller than or equal to 80 ms as anticipatory.

Density maps of eye positions over time

To visualize the temporal evolution of eye positions in our experiments, we computed movies of two-dimensional density maps for the different pretrial fixation times and each eye position of the time series recorded for each experiment. Based on a kernel density estimation via diffusion (Botev, Grotowski, & Kroese, 2010), we estimated density maps

for the first 2 s (after removal of the fixation cross) in each experiment. These movies are available as supplementary material under
http://jov.arvojournals.org/article.aspx?articleid=2661519.

3.3 Experiment 1

3.3.1 Methods

Participants

We recorded eye movements from 40 participants in Experiment 1 (34 female, 14–39 years old); 38 participants were recruited from the University of Potsdam and two from a nearby high school.

Procedure

In Experiment 1 the fixation cross was presented at the horizontal meridian 5.6° (256 pixels) away from the left or right border of the monitor. This position was chosen to reproduce the findings of a strongly reduced central fixation bias observed in an earlier study (Rothkegel et al., 2016), where participants experienced a pretrial fixation time of 1 s. A proportion of 20% of participants explored the image immediately after successful fixation without an additional pretrial fixation time (0 ms). This corresponds with the standard scene viewing paradigm. For all other participants the fixation cross remained on top of the image for a duration of 125 ms, 250 ms, 500 ms, or 1000 ms. Pretrial fixation time was used as a between-subject factor, i.e., each participant was tested with one of five pretrial fixation times. Figure 3.1 illustrates a representative trial with the starting position on the left side of the screen. Fixation Check 2 was nonexistent for participants with a 0-ms pretrial fixation time.

3.3.2 Results

Distance to center over time

In Experiment 1, the *DTC* initially decreased for all conditions (i.e., the CFB increased; see Fig. 3.2). There was a pronounced effect that mean fixation positions tended to be closer to the image center when participants were allowed to explore an image immediately after image onset, i.e., with a pretrial fixation time of 0 ms (black curves in Fig. 3.2a). Surprisingly, for the first four participants (Block 1) of this group the effect was visible throughout the whole observation time of 5 s. A second group of participants in the 0 ms condition (Block 2) did not replicate the stronger CFB through the whole observation

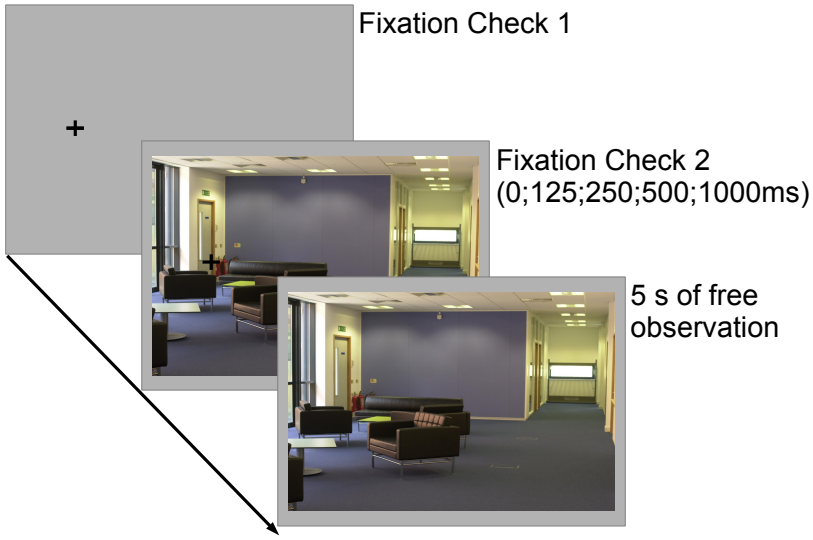

Figure 3.1: Schematic illustration of the experimental procedure of Experiment 1 with a starting position close to the left border of the screen. After a short fixation check of 200 ms (Fixation Check 1) the image is presented. A second fixation check between 0 and 1000 ms controls if participants move their eyes after image onset. After a successful second fixation check, participants are allowed to freely move their eyes.

a b

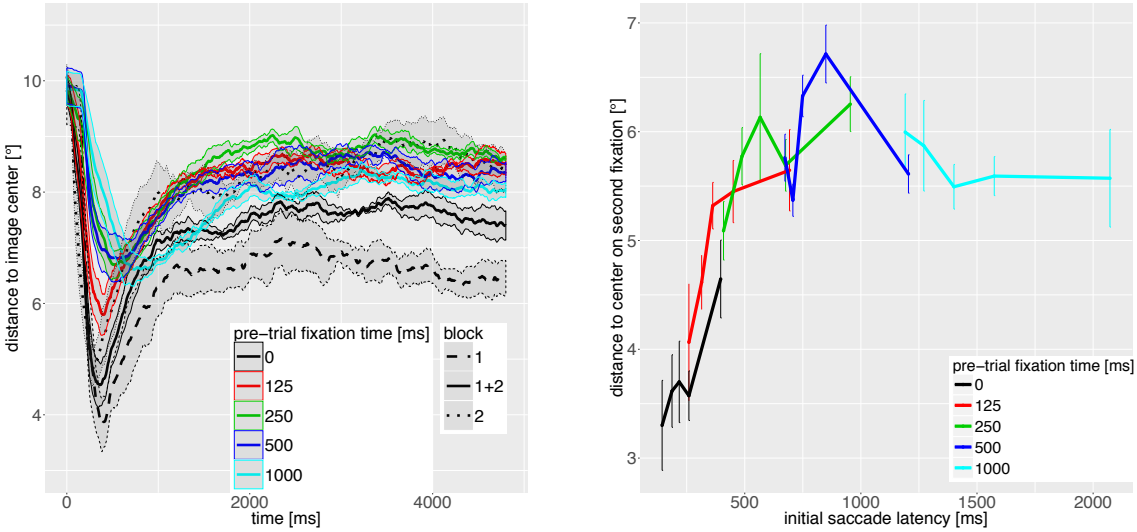

Figure 3.2: Experiment 1. a) Mean distance to center over time ($DTC(t)$) for the five different pretrial fixation times with starting positions close to the border of the screen. Confidence intervals indicate standard errors as described by Cousineau (2005). Block 1 represents participants 1-20, Block 2 participants 21-40 who were originally tested as a follow up experiment to consolidate the results. b) Mean distance to center of the second fixation as a result of initial saccade latency and pretrial fixation time. Bins represent quintiles of the saccade latency distribution. Errorbars are the standard error of the mean.

time. In addition, there was a gradual reduction of the CFB for pretrial fixation times from 125 ms to 250 ms (red and green curve). *DTC* for pretrial fixation times of 250 and 500 ms hardly differed (green vs. blue curve). The minimum for the pretrial fixation time of 1000 ms occurred later in time because of disproportionately long saccade latencies of the first saccade after a forced fixation on the fixation cross of 1000 ms (cyan curve).

Distance to Center on the second fixation

Figure 3.2b shows the influence of initial saccade latency on the mean DTC of the second fixation for the five pretrial fixation times. Each bin represents a quintile of the distribution of saccade latencies in each condition. A clear relation between DTC of the second fixation and latency of the initial saccade is visible for the pretrial fixation times of 0 ms and 125 ms. Overall, short saccade latencies led to a small average DTC (i.e., a strong initial CFB) whereas long latencies led to a larger average DTC (i.e., a less pronounced initial CFB).

Table 3.1 shows the output of the LMM for Experiment 1. The DTC for a pretrial fixation time of 0 ms is significantly lower than the average DTC and for a pretrial fixation time of 500 ms it is significantly higher. The initial saccade latency is highly significant regardless of the pretrial fixation time. This means that a saccade immediately after the sudden image onset led to a stronger CFB in this experiment. The model also shows that an interaction between saccade latency and pretrial fixation time exists. If participants are allowed to move their eyes directly after image onset (pretrial fixation of 0 ms), the influence of saccade latency is significantly higher than on average (see saccade latency $\times$ 0 ms). If pretrial fixation time is as long as 500 ms, the influence of saccade latency is significantly weaker than on average (see saccade latency $\times$ 500 ms). This interaction suggests that after a certain threshold time is reached, the influence of increasing saccade latency disappears.

3.3.3 Discussion

Experiment 1 led to a reduction of the CFB on the initial saccade target for all pretrial fixation times of 125 ms and more during scene perception from extreme starting positions (Fig. 3.2a). A pretrial fixation time of 125 ms produced an intermediate CFB, whereas longer pretrial fixation times produced an asymptotic behavior. With a pretrial fixation time of 0 ms the DTC was smaller throughout almost the whole observation time of 5 s for the first group of participants. However, this effect was not replicated in a retest with 20 new participants. The early effect of the CFB did not differ in the two groups of participants. The CFB of the second fixation did strongly depend on the latency of the initial saccade (Fig. 3.2b). Thus, the early differences between pretrial fixation times in

Table 3.1: Output of LMM for Experiment 1

Fixed Effect	Estimate	SE	t	
(Intercept)	1.856	0.079	23.546	***
0 ms	-0.925	0.151	-6.106	***
125 ms	-0.148	0.140	-1.053	
250 ms	0.185	0.140	1.320	
500 ms	0.496	0.136	3.660	***
saccade latency	0.751	0.108	6.951	***
saccade latency x 0 ms	1.685	0.339	4.976	***
saccade latency x 125 ms	0.245	0.208	1.178	
saccade latency x 250 ms	-0.210	0.175	-1.199	
saccade latency x 500 ms	-0.886	0.155	-5.726	***
Random effects variance: Subjects		0.1498		
Random effects variance: Images		0.1477		
Log-Likelihood		-7135.53		
Deviance		14271.07		
AIC		14297.07		
BIC		14380.64		
N		4575		

$^{*}p < .05,\ ^{**}p < .01,\ ^{***}p < .001$

Figure 3.2a are driven by differences in the distribution of initial saccade latencies.

These results replicated our earlier findings of a reduced CFB during scene perception by introducing a non-zero pretrial fixation time (Rothkegel et al., 2016). A delay of 125 ms was sufficient to achieve a considerable reduction and after a delay of 250 ms the minima of DTC curves only differed marginally. In addition, our results suggest that the most important mediating factor of the CFB was the latency of the first saccadic response. Saccades with brief saccade latencies were on average directed more strongly toward the center than saccades with long saccade latencies.

3.4 Experiment 2

To assure that our results from Experiment 1 were not mainly induced by the extreme starting positions we conducted another experiment with starting positions closer to the image center.

3.4.1 Methods

Participants

We recorded eye movements from 20 participants for Experiment 2 (17 female; 14-28 years old). Nineteen subjects were recruited from the University of Potsdam and one from a

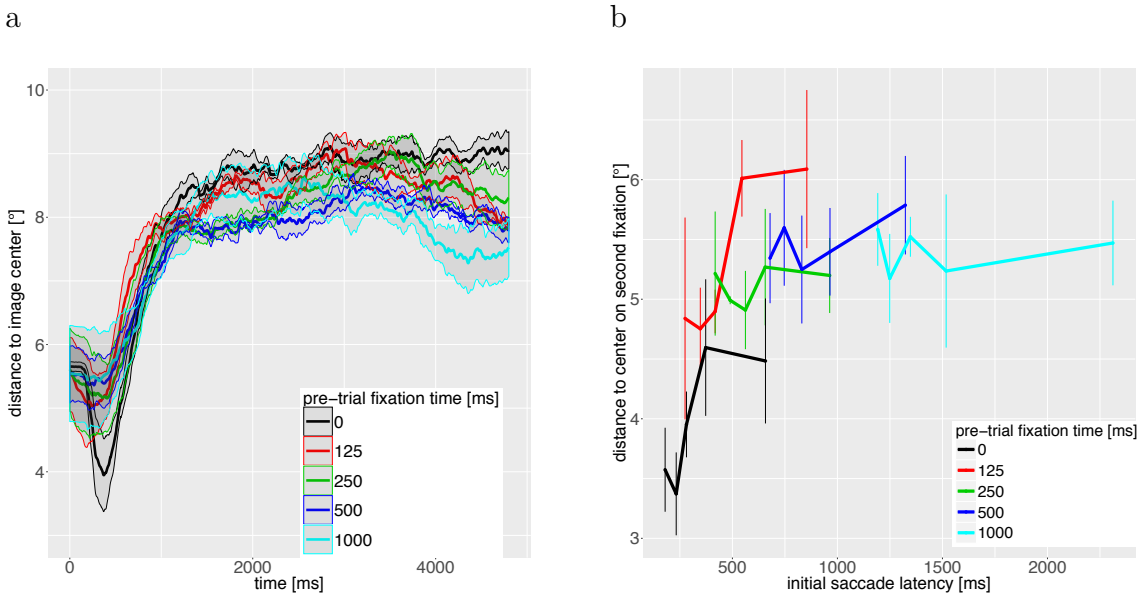

Figure 3.3: Experiment 2. a) Mean distance to center over time ($DTC(t)$) for the five different pretrial fixation times with starting positions on a donut shaped ring around the image center. Confidence intervals indicate standard errors as described by Cousineau (2005). b) Mean distance to center of the second fixation as a result of initial saccade latency and pretrial fixation time. Bins represent quintiles of the saccade latency distribution. Errorbars are the standard error of the mean.

nearby high school.

Procedure

Experiment 2 was similar to Experiment 1 except that the fixation cross was presented on a donut-shaped ring with a distance of 2.6° to 7.8° (100-300 pixels) to the center. We used this donut-shaped ring to obtain intermediate starting positions neither too close nor too far away from the center so that fixations could be directed both towards and away from the center. In addition, the donut-shaped ring of starting positions made the initial starting position less predictable. This setup differed slightly from the experiment conducted by Tatler (2007) where the initial starting position was randomly chosen from a circle (fixed radius) around the image center.

3.4.2 Results

Distance to center over time

In Experiment 2, where the starting positions were located on a ring around the image center, the eyes moved initially even further towards the image center in the 0 ms pretrial condition (black curve in Fig. 3.3a was the only curve with a pronounced negative slope

Table 3.2: Output of LMM for Experiment 2

Fixed Effect	Estimate	SE	t	
(Intercept)	2.167	0.094	22.965	***
0 ms	-0.899	0.186	-4.829	***
125 ms	-0.010	0.177	-0.054	
250 ms	0.228	0.174	1.308	
500 ms	0.255	0.177	1.439	
saccade latency	0.468	0.110	4.258	***
saccade latency x 0 ms	1.126	0.291	3.870	***
saccade latency x 125 ms	0.327	0.229	1.430	
saccade latency x 250 ms	-0.541	0.204	-2.656	**
saccade latency x 500 ms	-0.235	0.208	-1.132	
Random effects variance: Subjects		0.1112		
Random effects variance: Images		0.1333		
Log-Likelihood		−3599.60		
Deviance		7199.20		
AIC		7225.20		
BIC		7299.56		
N		2253		

$^{*}p < .05,\ ^{**}p < .01,\ ^{***}p < .001$

in the beginning). A difference in *DTC* was visible until about 600 ms after offset of the fixation marker. Later during the trial, the curves converged for all pretrial conditions and reached a stable DTC for the rest of the trial. Qualitatively, we also observed a small initial difference in DTC between short pretrial fixation times of 125 ms and 250 ms and pretrial fixation times of 500 ms and 1000 ms.

Distance to center of the second fixation

As in Experiment 1, we found a strong influence of the latency of the first saccade on the DTC of the second fixation for small pretrial fixation times (Fig. 3.3b). The results of the linear mixed model in Experiment 2 (Tab. 3.2) were similar to Experiment 1. The most important results are the significantly lower DTC of the 0 ms pretrial fixation time compared with the average and the significant increase in DTC for higher saccade latencies. As in Experiment 1 an interaction between saccade latency and pretrial fixation time is visible. This is especially true for the 0 ms condition, where the influence of saccade latency significantly increases compared with the average influence. In Experiment 2 the only significant decrease in saccade latency influence is visible for a pretrial fixation time of 250 ms. Overall direction of the influence (increasing influence of saccade latency for pretrial fixation times of 0 ms and 125 ms vs. decreasing influence for pretrial fixation times of 250 ms and 500 ms) is the same as in Experiment 1.

3.4.3 Discussion

If the starting position was close to the image center all pretrial fixation times of 125 ms or longer (Fig. 3.3a) led to a reduction of the CFB on early fixations. After around 600 ms this influence disappeared. Furthermore, a clear relation between latency of the first saccade and the CFB of the second fixation was visible (Fig. 3.3b). Thus, the results replicated our observations from Experiment 1 and demonstrated that a reduced CFB was not exclusively generated by the extreme starting positions used in Experiment 1.

3.5 Experiment 3

The results from Experiment 1 and 2 showed that a pretrial fixation time of 125 ms was enough to reduce the central fixation bias on early fixations. The difference of the CFB between pretrial fixation times larger than 125 ms was relatively small. To investigate the minimum pretrial fixation time for a substantial CFB reduction, we conducted a third experiment with pretrial fixation times ranging from 0 to 125 ms in six equidistant steps. We changed the between-subject design of pretrial fixation time to a within-subject design to reduce the influence of individual participants (cf., Exp. 1). Hence, every participant was tested with all pretrial fixation times. Because effects were maximal in the first experiment we used the same extreme starting positions as in Experiment 1.

3.5.1 Methods

Participants

We recorded eye movements from 24 participants for Experiment 3 (20 female; 20–29 years old). All participants were recruited from the University of Potsdam.

Procedure

In Experiment 3, participants experienced pretrial fixation times between 0 and 125 ms in steps of 25 ms (0, 25, 50, 75, 100, 125 ms). Each of the six pretrial fixation times was presented in a block of 20 images, pseudorandomized across participants. Note that the experiment was tested with a different setup (monitor, eye tracker, etc.; see General methods section for details). Thus, the absolute value of DTC is not directly comparable between Experiment 3 and the remaining experiments.

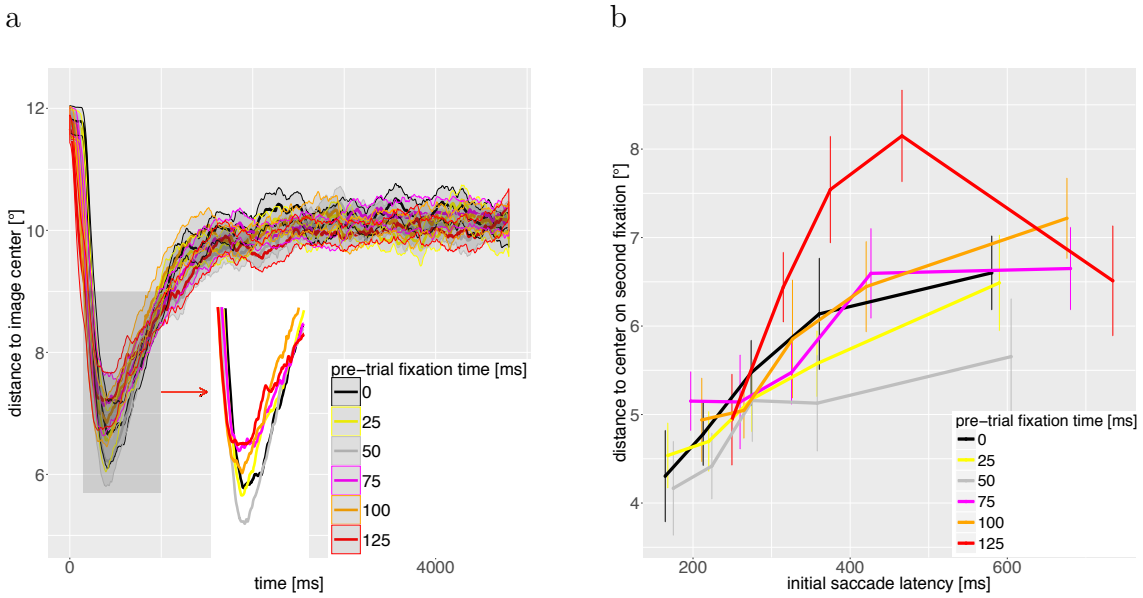

Figure 3.4: Experiment 3. a) Mean distance to center over time ($DTC(t)$) for the six different pretrial fixation times with starting positions close to the left and right border. Confidence intervals indicate standard errors as described by Cousineau (2005). b) Mean distance to center of the second fixation as a result of initial saccade latency and pretrial fixation time. Bins represent quintiles of the saccade latency distribution. Errorbars are the standard error of the mean.

3.5.2 Results

Distance to center over time

As in Experiment 1 and 2 the eyes initially moved towards the center for all pretrial fixation times (Fig. 3.4a). The difference between pretrial conditions was not as clearly visible as in previous experiments. Even the difference between the 0- and the 125-ms condition was relatively small. The smaller difference was probably due to the blocked design where pretrial fixation times changed after 20 trials during the experiment for each participant. Nonetheless, curves with a pretrial fixation time smaller than or equal to 50 ms had smaller minima than the ones with pretrial fixation times larger than 50 ms (see inset in Fig. 3.4a).

Distance to center of the second fixation

The influence of the first saccade latency on the distance to center of the second fixation is clearly visible in Figure 3.4b. The influence seemed even clearer than in previous experiments. However, the range of the distance to center values was larger in this experiment as a result of the increased magnitude of the image in visual degree. Saccade latencies were more homogeneous in Experiment 3. The difference of mean saccade latencies (pre-

Table 3.3: Output of LMM for Experiment 3

Fixed Effect	Estimate	SE	t	
(Intercept)	2.187	0.110	19.832	***
0 ms	0.056	0.076	0.731	
25 ms	-0.096	0.086	-1.118	
50 ms	-0.069	0.072	-0.949	
75 ms	0.055	0.081	0.673	
100 ms	-0.015	0.078	-0.190	
saccade latency	1.056	0.110	9.564	***
saccade latency x 0 ms	-0.334	0.201	-1.662	
saccade latency x 25 ms	0.161	0.249	0.645	
saccade latency x 50 ms	0.071	0.208	0.340	
saccade latency x 75 ms	-0.093	0.227	-0.409	
saccade latency x 100 ms	0.153	0.226	0.678	
Random effects variance: Subjects		0.2308		
Random effects variance: Images		0.1359		
Log-Likelihood		−3990.68		
Deviance		7981.35		
AIC		8011.35		
BIC		8099.75		
N		2679		

$^{*}p < .05,\ ^{**}p < .01,\ ^{***}p < .001$

trial fixation time + saccade latency after removal of the fixation marker) between the 0- and 125- ms condition was much smaller (57 ms) than in Experiments 1 (154 ms) and 2 (138 ms).

A linear mixed model for Experiment 3 showed that DTC of the second fixation did not show an independent influence of pretrial fixation time (Tab. 3.3). However, we replicated a significant influence of the first saccade latency on DTC of the second fixation. Shorter saccade latencies led to fixations closer to the center of an image. An interaction between pretrial fixation time and saccade latency was not observed.

Distributions of saccade latencies in Experiment 3 were rather similar between different pretrial fixation times. However, there was a difference between the three lowest pretrial fixation times (mean saccade-latencies of 315, 320, and 321 ms) compared with the three longer pretrial fixation times (mean saccade-latencies of 365, 352, and 371 ms). Thus somewhere around 75 ms seems to be the lowest pretrial fixation time to influence further viewing behavior.

3.5.3 Discussion

Experiment 3 was conducted to investigate the minimum pretrial fixation time necessary for a reduction of the early central fixation bias. All pretrial conditions showed a similar

behavior with a tendency of an early CFB as measured by the DTC. We observed the weakest DTC effect for pretrial fixation times of 125 ms (inset in Fig. 3.4a). Pretrial fixation times equal to or smaller than 50 ms generated fixation positions closest to the image center. Differences in DTC could be explained by the influence of the first saccade latency on the selection of the second fixation location (Fig. 3.4b). Thus, saccade latencies are the most important factor modulating the CFB. A post-hoc analysis revealed that saccade latencies were only affected in conditions with pretrial fixation times larger than 50 ms. This is in line with previous research that the shortest image preview to influence further eye movement behavior in visual search lies between 50 and 75 ms (Võ & Henderson, 2010). We conclude that a minimum pretrial fixation time of around 75 ms is needed to prolong saccade latencies in order to reduce the CFB in scene viewing.

3.6 Experiment 4

In Experiment 4, participants started exploration at the center of the screen. This starting position was chosen to quantify the influence of pretrial fixation times in a standard scene viewing paradigm.

3.6.1 Methods

Participants

In this experiment we recorded eye movements from 10 participants (three male; 18–36 years old). All were recruited from the University of Potsdam.

Procedure

Experiment 4 followed the same procedure as the preceding experiments but participants started observation in the center of the screen. We tested pretrial fixation times of 0, 125, and 250 ms since we observed only subtle changes of results for longer pretrial fixation times in Experiments 1 and 2. As in Experiment 3, we used a within-subject design for the three different pretrial fixation times such that participants viewed blocks of 40 images for each pretrial fixation time.

3.6.2 Results

Distance to center over time

Contrary to the first experiments initial gaze positions could only move away from the image center with central starting positions in Experiment 4 (Fig. 3.5a). Therefore, DTC

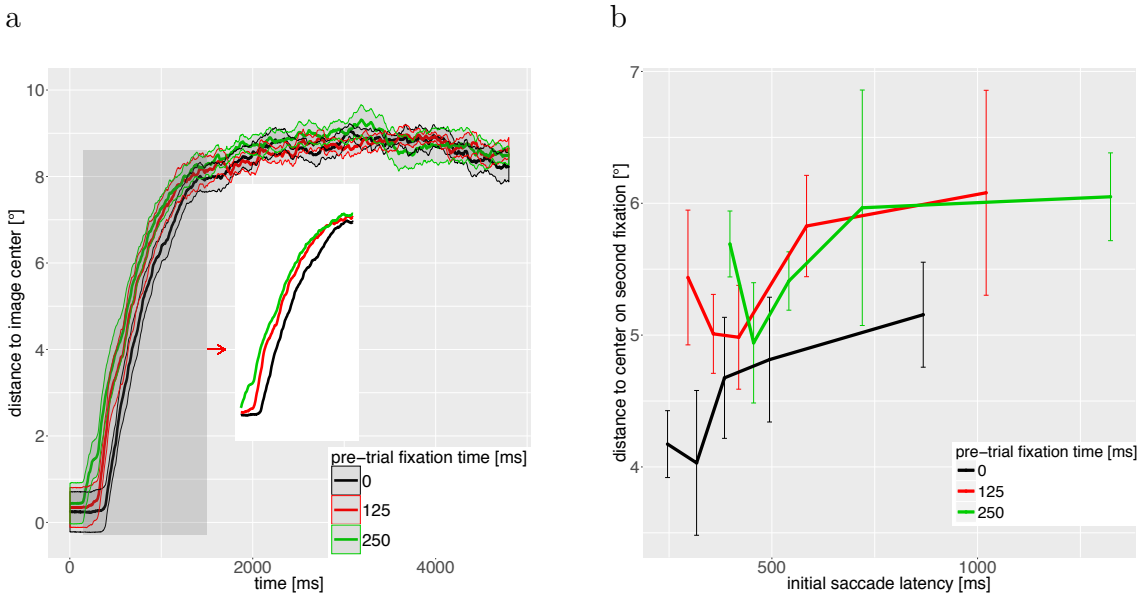

Figure 3.5: Experiment 4. a) Mean distance to center over time ($DTC(t)$) for the three different pretrial fixation times with starting positions in the center of the image. Confidence intervals indicate standard errors as described by Cousineau (2005). b) Mean distance to center of the second fixation as a result of initial saccade latency and pretrial fixation time. Bins represent quintiles of the saccade latency distribution. Errorbars are the standard error of the mean.

gradually increased until it reached an asymptote. Between pretrial conditions, DTC differed with respect to the point in time, when curves started to monotonically increase (pretrial fixation times: 250 ms < 125 ms < 0 ms). Although pretrial fixation times were chosen to be equidistant, curves for 125-ms and 0-ms pretrial conditions (red and black curve) take longer to converge than curves for the 250-ms and 125-ms pretrial conditions (green and red curve; see inset Fig.3.5a). This demonstrated that pretrial fixation times of 125 ms or more reduce the CFB of early fixations even during scene viewing with central starting positions.

Distance to center of the second fixation

Latencies of the first saccade were longer in this experiment than in any of the other experiments. This observation is in line with results from face perception, where the initial fixation is longer when participants start exploring a face in the center (Arizpe, Kravitz, Yovel, & Baker, 2012). Due to the increased number of long initial saccade latencies, an influence of saccade latency on the second fixation location was not as clearly visible as in the previous experiments (Fig. 3.5b).

Results of a linear mixed model for Experiment 4 partially replicated the main results from Experiments 1–3. The DTC of the second fixation was significantly smaller for a

Table 3.4: Output of LMM for Experiment 4

Fixed Effect	Estimate	SE	t	
(Intercept)	2.858	0.157	18.194	***
0 ms	-0.309	0.093	-3.312	***
125 ms	0.155	0.083	1.877	
saccade latency	0.224	0.122	1.837	
saccade latency x 0 ms	0.454	0.174	2.611	**
saccade latency x 125 ms	-0.189	0.154	-1.222	
Random effects variance: Subjects		0.2607		
Random effects variance: Images		0.1845		
Log-Likelihood		−1877.00		
Deviance		3754.01		
AIC		3772.01		
BIC		3817.26		
N		1128		

$^{*}p < .05,\ ^{**}p < .01,\ ^{***}p < .001$

pretrial fixation time of 0 ms. The influence of saccade latency on distance to center of the second fixation did not reach a level of significance of 95% in Experiment 4. The direction of the influence was positive and nearly reached the level of significance. The fact that saccade latency was not a significant predictor is a result of the rather long latencies and a small number of participants. By removing initial saccade latencies of higher than 1 s (which normally are very rare) saccade latency becomes a significant predictor ($p < 0.03$). The interaction between saccade latency and pretrial times showed that the influence of saccade latency on DTC was, as observed in Experiments 1 and 2, significantly larger for a pretrial fixation time of 0 ms.

3.6.3 Discussion

In our last experiment we investigated the effect of pretrial fixation times on the CFB in a standard scene-viewing experiment where participants start exploration from the image center. As expected, DTC increased in all conditions continuously until it reached an asymptote. The point in time when DTC started to increase varied for different pretrial fixation times. We measured the earliest response for pretrial fixation times of 250 ms and the slowest response after no pretrial fixation times (0 ms). If we remove latencies of higher than 1 s we can replicate an influence of saccade latencies on DTC of the second fixation. In general, saccade latency seems to be a strong mediating factor of the CFB. In addition, we observed long initial saccade latencies when participants started at the image center. This is particularly worrying, because the first fixation is usually omitted from analyses in scene viewing experiments.

When comparing Experiment 4 to the remaining experiments, the CFB was strongest when participants started at the image center without pretrial fixation time (0 ms). Only after about 1 s DTC (and CFB) was comparable between experiments and pretrial conditions. Because most scene-viewing experiments last five seconds or less (c.f., data sets in MIT saliency benchmark; Bylinskii et al., 2015) a substantial proportion of fixations is biased towards the center during a standard scene viewing experiment. A combination of a non-zero pretrial fixation time and adjustments of the starting position will reduce the CFB and may help to better understand target selection during scene viewing. We will further comment on this issue in the general discussion.

3.7 Discussion of empirical results

In four scene-viewing experiments we have shown that by delaying the initial saccade relative to the sudden image onset the early central fixation bias was significantly reduced. Further analysis showed that the amount of early CFB is directly linked to the initial saccade latency. Figure 3.6 shows the influence of initial saccade latency on distance to image center for all four experiments combined. A clear increase of DTC is visible between 150 and 400 ms. Because initial saccade latencies above 400 ms do not show an influence, pretrial fixation times above 250 ms did not produce noteworthy effects. This also explains why in Experiment 4 the rather long saccade latencies were not a significant predictor for the CFB. We conclude our experiment by stating that the initial saccade latency is the dominant factor influencing the early central fixation bias in scene viewing. This leads to the assumption that the sudden image onset is involved in generating the early CFB.

3.8 Computational modeling of the central fixation bias

To test if the early CFB might result from default activation in the image center after a sudden onset that is replaced by a content driven activation over time, we simulated scanpaths generated by a computational model. For the simulations we used an extended version of the previously published SceneWalk model of saccade generation from our group (Engbert et al., 2015). Different to the original model with zero activation at the beginning of a trial, we decided to start each trial with higher activations in the center of an image than at the periphery (see Fig. 3.7a). The influence of this central starting activation declines with increasing saccade latency and is replaced by a more content driven activation (the empirical density map of the image multiplied with a Gaussian

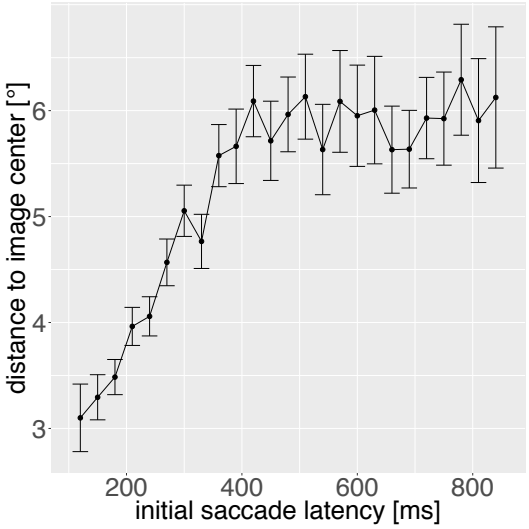

Figure 3.6: Influence of initial saccade latency on the distance to image center of the second fixation for all 4 experiments combined.

around the starting position; see Fig. 3.7b). This initial central activation represents the sudden image onset. We refer to this extended model as the SceneWalk StartMap model. A more detailed description of the model can be found in the Appendix.

Figure 3.7 shows the simulated fixations 1–4 of two trials with different pretrial conditions (0 ms vs. 1000 ms) of the SceneWalk StartMap model. The initial saccade latency in the first trial (Fig 3.7a) was very short ($t = 184\ ms$) and thus the first target selection map of the SceneWalk StartMap model is biased strongly toward the center. The activations on this map translate into probabilities for being "fixated" by the model. Thus trials with short initial saccade latencies produce many fixations close to the image center. The second trial (Fig 3.7b) had an initial saccade latency of 1484 ms (1000-ms pretrial fixation time + 484 ms after the fixation cross vanished) which is enough to replace the central activation map with the empirical density map of the image multiplied with a Gaussian around the starting position. After a long saccade latency, this map is roughly the same map as the original SceneWalk model without an explicit center bias produces and leads to mean fixation positions further away from the image center.

We simulated saccadic sequences from the SceneWalk StartMap model with the same starting positions, number of fixations, and fixation durations as observed empirically. The temporal evolution of the DTC of Experiment 1 for different pretrial fixation times for the SceneWalk StartMap model is shown in Figure 3.8a). The SceneWalk StartMap model took the initial saccade latency after image onset into account, which produced a qualitatively similar pattern for the different pretrial fixation times as seen in the data. The qualitative progression for most pretrial fixation times was similar to what was observed empirically. It is eminent though that the central fixation tendency produced by

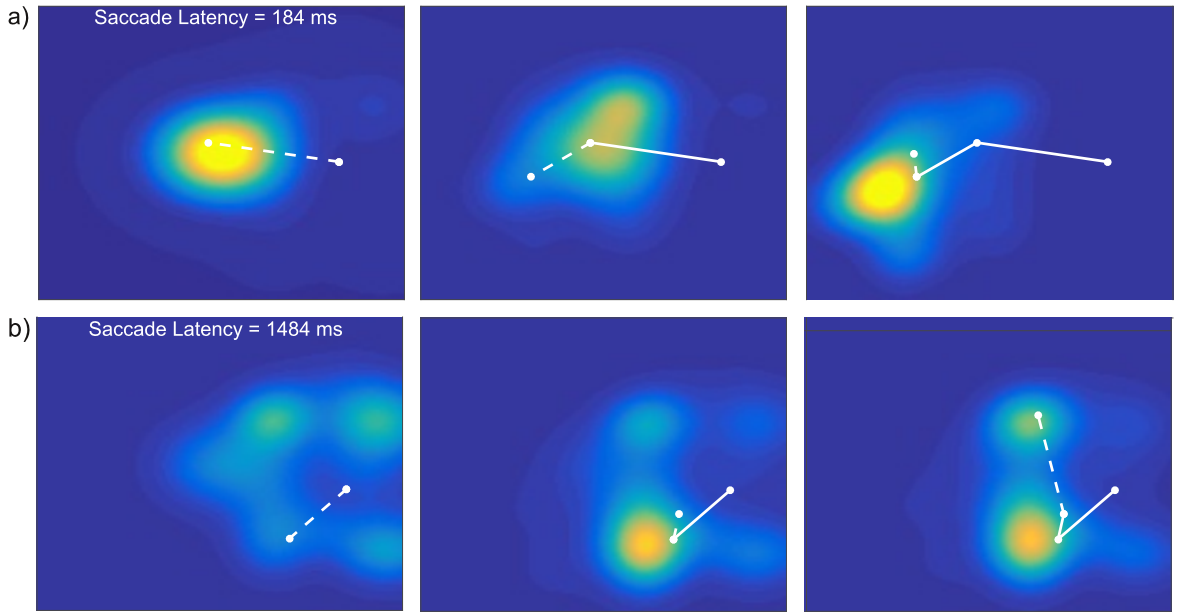

Figure 3.7: Simulated fixations 1–4 (left to right) of two trials on the same image. a) A pretrial fixation time of 0 ms and a saccade latency of 184 ms create an attention map for the first saccade target biased towards the center. This leads to fixations close to the center. b) A pretrial fixation time of 1000 ms and saccade latency of 1484 ms create an attention map for the first saccade target without a central bias. The initial attention map in this trial roughly represents the empirical density map of the image multiplied with a Gaussian around the starting position.

the model was too weak when compared with the data. This was probably a result of the method and the fixations used for the parameter estimation (see Appendix).

We also evaluated the relation between latencies of the first saccade and DTC of the second fixation (Fig. 3.8b). This influence was also visible in the SceneWalk StartMap model, because longer initial saccade latencies led to a less pronounced central activation map. The SceneWalk StartMap model produced a result pattern similar to the empirical data with a similar progression of lines and a differentiation between pretrial fixation times. However, the early CFB on the second fixation was too small in all experiments, i.e., the distance to center in all simulations was too large.

3.8.1 Discussion

Adjusting an existing model of saccade generation with an initial central activation map whose influence declines with increasing saccade latency can reasonably explain the central fixation bias. The SceneWalk StartMap model qualitatively replicated differences

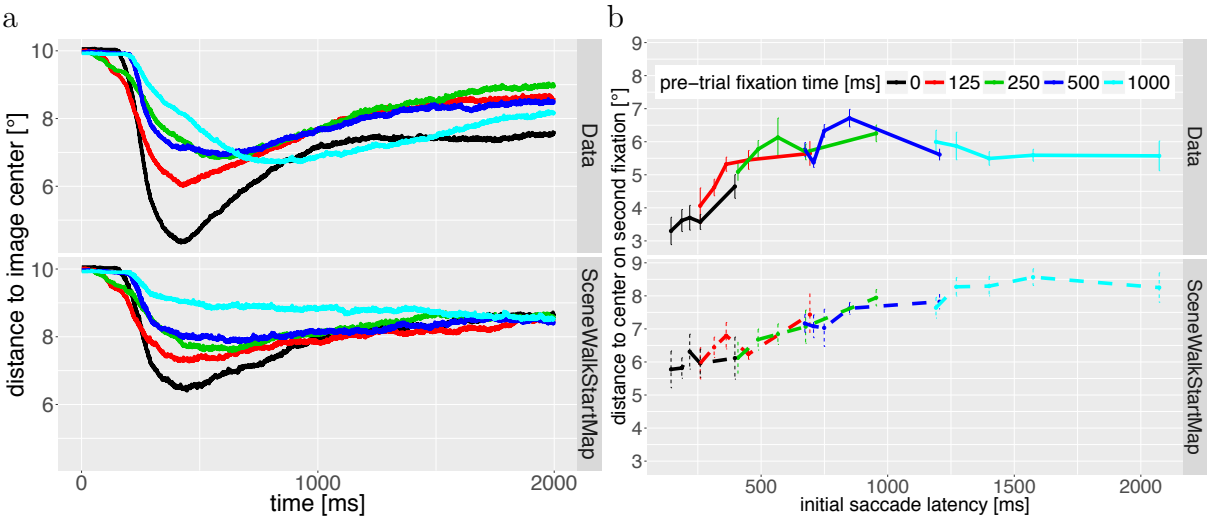

Figure 3.8: a) Distance to image center over time for the empirical data and the SceneWalk StartMap model for the different pretrial fixation times in Experiment 1. b) Influence of the initial saccade latency on the distance to image center on the second fixation for the empirical data and the SceneWalk StartMap model in Experiment 1.

in DTC curves between pretrial fixation times, and replicated saccade latency effects on DTC of the second fixation. However, the CFB from our simulations was too weak, which is probably a result from the methods used for estimating the parameters (see Appendix). Replicating our empirical findings was accomplished by assuming that a the central fixation bias is a result of a default activation in the center of a suddenly appearing stimulus, which is gradually replaced by a content-driven activation.

3.9 General discussion

During scene viewing the eyes have a strong tendency to fixate near the center of an image, which potentially masks other bottom-up and top-down effects of saccadic target selection. In a previous study (Rothkegel et al., 2016) with starting positions near the image border and an experimentally delayed first saccade after the onset of an image we observed a considerable reduction of the central fixation bias (CFB; Tatler, 2007). Here, we investigated this reduction in four scene-viewing experiments. We manipulated starting positions and the latency of the initial saccade. Different to the original scene-viewing paradigm, where participants start exploration immediately after image onset, we delayed the initial saccadic response by instructing participants to start exploration only after disappearance of a fixation marker. As a measure of the central fixation bias we computed the distance to center (DTC) of the eyes over time. In all experiments the disappearance of a fixation marker 125 ms after image onset led to an early reduction of the CFB in comparison to trials where the fixation marker disappeared simultaneously to

the image onset (original scene-viewing paradigm). The earliest pretrial fixation time to produce an influence was measured at 75 ms (see Exp. 3). The reduction of the CFB was particularly pronounced in experiments with pretrial fixation time as a between-subject factor (Exp. 1 and 2). A reduction of the CFB was even visible when participants started observation at the center of an image (Exp. 4). The distance to center of the second fixation was well predicted by the latency of the initial saccade (time from image onset) across experiments. Short saccade latencies led to a strong bias toward the center whereas longer saccade latencies were less systematically directed toward the center. Hence, the latency of the initial response seemed to primarily account for the observed differences of the CFB.

Previous studies have shown that it takes 90 ms on average for the visual input to reach the cortical areas (V. P. Clark, Fan, & Hillyard, 1995; Di Russo, Martínez, Sereno, Pitzalis, & Hillyard, 2002) and at least another 60 ms to execute an already programmed saccade (e.g., Findlay & Harris, 1984; Ludwig, Mildinhall, & Gilchrist, 2007). Thus, to plan a saccade to an image dependent location the latency has to be at least 150 ms. All pretrial fixation times smaller than 125 ms contained trials with initial saccades latencies below 150 ms. Thus implementing a pretrial fixation time of 125 ms and more removed all saccades, which could not have been the result of an image-specific target selection. Additionally, our results and model simulations have shown that the central fixation bias gradually decreases for latencies from 150 ms to 400 ms. Thus we propose a delay somewhere between 125 ms and 250 ms for a strong and reliable reduction of the CFB.

Our findings are in agreement with the note communicated earlier that a sudden image onset during scene viewing represents an artificial laboratory situation and may cause unnatural saccadic behavior ('t Hart et al., 2009; Tatler et al., 2011). However, the sudden image onset seems to primarily affect the tendency of the first saccade to move the eyes toward the center of an image. Due to the dependence of fixation locations (Engbert et al., 2015), subsequent fixations are then also more likely located near the center. Of the two explanations for the CFB proposed by Tatler (2007), one can be excluded from our results. If the image center is the strategically optimal position to start inspection of the image, regardless of the content and previous gist extraction, the central fixation bias would not decrease due to prolonged initial saccade latency. The other remaining possibility of the central fixation bias was that by fixating the center of the image the amount of information or gist being extracted is maximized. This explanation cannot be ruled out due to our results. However, if participants are forced to extract the scenes' gist from another position, they do not necessarily look at the center for further information extraction.

We propose another explanation for the early central fixation bias. Our results have shown that the sudden image onset is a dominant contributor to the persistent early

central fixation bias. Previous research has shown that the sudden appearance of a new stimulus captures attention and attracts eye movements, even if it is completely task irrelevant (Theeuwes et al., 1998). If, however, this suddenly appearing stimulus appears during a fixation when no saccade is being programmed it does not guide gaze irrespective of the task (Tse, Sheinberg, & Logothetis, 2002). These results can be transferred to our experiment in the following way: The sudden luminance change on the monitor when an image is displayed can be treated as a large object (the image) suddenly appearing. When looking at an object, the eyes usually try to land in the center (Nuthmann & Henderson, 2010) and suddenly appearing objects are being fixated close to the center (Richards & Kaufman, 1969; Kowler & Blaser, 1995). Thus, if an image suddenly appears and a saccade is planned in parallel, this saccade is a reflexive, stimulus driven saccade towards the appearing stimulus executed via a subcortical path (Ottes, Van Gisbergen, & Eggermont, 1985; Munoz & Everling, 2004). If some time passes after the sudden onset, before the saccade is executed, a saccade can be planed via a cortical path (Ottes et al., 1985; Munoz & Everling, 2004) targeting a location defined by the image content.

These hypotheses were used to extend a recently published model of saccade generation (SceneWalk model; Engbert et al., 2015; Schütt et al., 2017). To generate a strong early CFB, we needed to assume that the sudden image onset led to a strong central activation at the beginning of a trial, which declines with increasing saccade latency. The model was able to qualitatively reproduce the CFB and the relation between saccade latency and the distance to center of the second fixation. However, in its current form the model underestimated both effects. These model simulations show that by only incorporating a central fixation bias, which depends on initial saccade latency, we were able to reproduce the progression of the early central fixation bias.

Computational models that aim at predicting the allocation of visual attention on an image are based on the extraction of image features (Itti et al., 1998; Borji & Itti, 2013) and top-down cognitive processes (Navalpakkam, Arbib, & Itti, 2005; Cerf et al., 2008). These models are evaluated by comparing human fixations with a weighted distribution of different influences (Bylinskii et al., 2015; Borji & Itti, 2013; Le Meur & Baccino, 2013; Borji, Cheng, Jiang, & Li, 2015). Although bottom-up and top-down influences as well as a combination of the two can predict human fixations (Bylinskii et al., 2015), the CFB is a strong predictor that improves goodness-of-fit more than any other single feature (Judd et al., 2009; Bylinskii et al., 2015). Thus, saliency models are usually compared with the CFB as a baseline (e.g., Wilming, Betz, Kietzmann, & König, 2011; Clarke & Tatler, 2014; Bruce, Wloka, Frosst, Rahman, & Tsotsos, 2015) and rely heavily on the implementation of a CFB for a good performance (Kümmerer et al., 2015). Because the early CFB during scene viewing seems to be an automated, stereotyped response of the saccadic system to a sudden image onset, it masks bottom-up and top-down factors of saccade target selection

and its strength critically depends on the duration of a trial since it primarily affects early fixations. Therefore, a reduction of the CFB during scene viewing, as generated by our paradigm, provides a better understanding of target selection and a more rigorous test of visual attention models than the original scene-viewing paradigm. At the minimum, the latency of the first saccade needs to be taken into account, because it strongly influences subsequent viewing behavior.

Our results imply to use a modified version of the scene-viewing paradigm to study bottom-up and top-down processes of target selection beyond the CFB. To minimize the influence of the sudden image onset, we suggest use of a fixation marker that disappears between 125 and 250 ms after image onset. In addition, due to the dependence of successive fixations, scene exploration should not exclusively start near the image center. Instead initial fixations (fixation markers) should be evenly distributed across the entire image or even with a preference toward the periphery. Central parts of the image will be fixated when the eyes move toward the other side of an image. Finally, sudden onsets of stimuli are often used in other laboratory tasks as well (e.g., visual search or face perception). To what extent our results generalize to other domains remains an open question but an early initial CFB might also bias initial fixations in these tasks.

3.9.1 Conclusion

Delaying the first saccadic response relative to image onset reduced the central fixation bias, which is most pronounced during early fixations. The latency of the first saccade after image onset was the main predictor for the distance to image center of the second fixation in all four experiments relatively independent of the time we enforced. The results suggest that the early central fixation bias is a result of default saccades as a response to a sudden image onset. Our results suggest use of modified version of the scene-viewing paradigm to better understand saccade target selection beyond the central fixation bias.

3.10 Appendix

3.10.1 SceneWalk Model

For our model simulations we took the existing SceneWalk model of saccade generation (Engbert et al., 2015) and extended it to model the early central fixation bias. The SceneWalk model proposes that eye movements are driven by two different time-dependent neural activation maps. An attention map reflects the attentional allocation on the given scene for a specific fixation position. To compute the attention map, first an intermediate map is computed by multiplying a two dimensional Gaussian distribution centered at

the current fixation position with the empirical saliency map of the image to reflect the reduced processing in the periphery. The influence of attention maps from previous fixations declines over time and thus the previous attention map is increasingly replaced by the map of the new fixation. A second map, the fixation map, memorizes previous fixations and tags visited fixations locations, making them less probable to be fixated again shortly afterwards. Thus, this map serves as an *inhibition of return* mechanism (Itti & Koch, 2001; Klein, 2000). The mechanism to control the dynamics of inhibition, i.e., the fixation map, is equivalent to the mechanism used for the attention map. The attention and inhibition maps prior to the first fixation are set to zero. After computation of the two maps for the current fixation position and duration, they are combined by subtracting the fixation map from the attention map to a target map. After the maps are combined, a target is chosen proportional to the relative activations (Luce, 1959) of the target map. Thus, positions where the fixation map is high whereas the attention map is low are rarely fixated and vice versa. For the interested reader the complete architecture of the model can be found in (Engbert et al., 2015) and a newer version in (Schütt et al., 2017).

3.10.2 SceneWalk StartMap Model

Since the original SceneWalk model was not intended to produce an early CFB, we developed a modified version of the original model which takes the sudden image onset during scene perception into account. We made two changes.

First, different than in the original SceneWalk model with zero activation across the entire attention map at the beginning of a trial, we used an attention map with higher activations near the center of an image than at the periphery (see Fig. 3.7a). This was motivated by the sudden image onset that may lead to an initial prioritization of central locations. This activation was a two dimensional Gaussian centered at the image center with two different standard deviations for the horizontal and vertical dimension (σ_x and σ_y). This initial attention map was normalized to a sum of 1.

Second, we realized that the decay of the attention map was too fast during the initial fixation. Therefore, we estimated a new parameter ρ_2 that specified the rate of decay during the initial fixation. For all other fixations we used the same decay parameter as during the original simulations (Engbert et al., 2015).

The default central activation maps transition into the attention map before the first saccade is computed as

$$a(t) = \phi \cdot A_{i,j}(t) + e^{(-t \cdot \rho_2)} \cdot (a(t) - \phi \cdot A_{i,j}(t)), \tag{3.2}$$

where $a(t)$ is the attention map at time (t) and $A_{i,j} \cdot \phi$ is the empirical density map multiplied with a Gaussian around the starting position i, j. The new decay parameter

ρ_2 controls the speed with which the initial central activation map is replaced. Thus with increasing saccade latency (increasing t) the initial central activation map (i.e. $a(0)$) is gradually replaced by the empirical saliency map multiplied with a Gaussian around the starting position ($\phi \cdot A_{i,j}$)

To estimate the parameters for the SceneWalk StartMap model we used a standard optimization algorithm (fminsearch) implemented in MATLAB (MATLAB, 2015) to obtain the parameters with maximum likelihood (Bickel & Doksum, 1977; Schütt et al., 2017) of fixations 2–4 of half of the participants (Exp. 1–4: $N = 20/10/12/5$) and a quarter of the images ($N = 30$). We estimated parameters from the second to fourth fixation only for efficiency reasons and since DTC curves reached a stable value for later fixations.

The horizontal standard deviation σ_x of the initial center map was estimated at values of $3.5°, 1.8°$ and $3.9°$ for Experiments 1–3. The vertical standard deviation σ_y for Experiment 1–3 was estimated at $2.3°, 2.3°$ and $2.4°$ and the decay parameters ρ_2 for the first three experiments were estimated at 1.11, 3.72 and 1.49. The parameters estimated for Experiment 4 were very large with $\sigma_x = 136.0°$, $\sigma_y = 4.2°$ and $\rho_2 = 310$. This resulted in small initial differences in activations between center and periphery for simulations of Experiment 4 and was similar to the constant activations in the original model. The reason for this behavior arises from the architecture of the model. Since activations in the attention map rise near fixation, central activations are prioritized initially when participants start to explore a scene near the image center.

Chapter 4

Searchers adjust their eye-movement dynamics to the target characteristics in natural scenes

Lars O. M. Rothkegel[1*], Heiko H. Schütt[1,2*], Hans A. Trukenbrod[1],

Felix A. Wichmann[2–4], and Ralf Engbert[1]

*These authors contributed equally to this work

[1]University of Potsdam, Germany

[2]Eberhard Karls University Tübingen, Germany

[3]Bernstein Center for Computational Neuroscience Tübingen, Germany

[4]Max Planck Institute for Intelligent Systems, Tübingen, Germany

Running head: Visual search

This is an earlier draft of a manuscript published 2019 in

Scientific Reports, Volume 9, Number 135. doi: 10.1038/s41598-018-37548-w

Abstract

When searching a target in a natural scene, both the target's visual properties and similarity to the background influence whether and how fast humans are able to find it. So far, it was unclear whether searchers adjust the dynamics of their eye movements (e.g., fixation durations, saccade amplitudes) to the target they search for. In our experiment participants searched natural scenes for six artificial targets with different spatial frequency content throughout eight consecutive sessions. High spatial frequency targets led to smaller saccade amplitudes and shorter fixation durations than low spatial frequency targets if target identity was known. If a saccade was programmed in the same direction as the previous saccade, fixation durations and successive saccade amplitudes were not influenced by target type. Visual saliency and empirical fixation density at the endpoints of saccades which maintain direction were comparatively low, indicating that these saccades were less selective. Our results demonstrate that searchers adjust their eye-movement dynamics to the search target efficiently, since low-spatial frequencies are visible farther into the periphery than high-spatial frequencies. We interpret the saccade direction specificity of our effects as an underlying separation into a default scanning mechanism and a selective, target-dependent mechanism.

4.1 Introduction

One of the most important everyday tasks of our visual system is to search for a specific target. Whether the task is to find a fruit amongst leaves, detect a dangerous animal or find relatives in a crowd of people, visual search has always been essential for survival. How the brain performs visual search tasks has been subject to a vast amount of research and, consequently, a number of comprehensive theories have been proposed (Treisman & Gelade, 1980; Wolfe, 1994; Duncan & Humphreys, 1989). However, most studies concerning visual search have been conducted on so-called search arrays, where targets and distractors are presented on a homogeneous background. While results from these highly controlled studies are very useful for understanding the basic nature of visual search, many do not take eye movements into account, although eye movements play an important role in real world search behavior (Findlay & Gilchrist, 2003; Malcolm & Henderson, 2009; Hulleman & Olivers, 2015).

When searching on a complex background, saccades—fast ballistic eye movements—are executed about three to four times per second to increase the probability of finding a target. It has been shown in many studies that the search target strongly influences saccade target selection of eye movements when searching through natural scenes. Object-scene consistency (Loftus & Mackworth, 1978; Henderson et al., 1999; T. H. Cornelissen & Võ, 2017), scene context (Torralba, 2003; Neider & Zelinsky, 2006) as well as low-level features (Hwang et al., 2009) of the target influence where observers fixate. Thus, a top-down search template of the target appears to guide gaze during scene exploration (Wolfe, 1994; Hwang et al., 2009). Correlations between the visual properties of target-related search templates and fixated image patches exist, but do not completely explain eye-movement behavior in visual search on complex backgrounds. Najemnik and Geisler (Najemnik & Geisler, 2005, 2008) showed that human observers do not simply move their eyes to positions which maximally resemble the target but rather apply a strategy which takes the visual degradation towards retinal periphery into account. They argue that observers sample as much relevant information as possible with a minimal number of eye movements, which they call the optimal eye movement strategy in visual search. Thus, it seems useful for the visual system to adapt eye-movement strategies according to the target's visibility in the periphery. Target visibility depends on retinal eccentricity (Meinecke, 1989) and its interaction with many factors such as spatial frequency (Pointer & Hess, 1989) and contrast (Campbell & Robson, 1968; Robson & Graham, 1981).

To investigate whether target features not only influence where participants look at (fixation locations) but also how they search (saccade amplitudes and fixation durations), we let participants search natural scenes for artificial targets with different low-level features. Although one might suspect that different targets lead to different saccade ampli-

tudes and fixation durations, to our knowledge no one has yet provided empirical evidence to answer this question. It is rather important for models of eye movement control to know whether, how fast, and how accurately human observers change search strategies contingent on the target they search for. To explicitly compare targets of different spatial frequency on various backgrounds, we used artificial targets instead of real-world objects in this study. Furthermore, we used scenes instead of plain backgrounds because (i) we are interested in real-world search behavior and not search on highly controlled arrays and (ii) to gain knowledge to improve dynamical models of saccade generation in natural scenes (Engbert et al., 2015; Schütt et al., 2017).

In our study, observers searched in each of 8 consecutive sessions for 6 targets of varying spatial frequency content and, in the case of high-spatial frequency, orientation (vertical and/or horizontal; see Fig. 4.1). Each session contained one block per target. Each block consisted of one repetition of the same 25 images. Target type was specified in advance to each block, to provide a search template. In one session (Session 7), targets were chosen randomly for each trial and target type was unknown prior to a trial.

If dynamical aspects of eye movements are indeed adapted to the search target in a useful way, saccade amplitudes should be larger during search for low-spatial frequency targets, since low-spatial frequencies can be detected further into the periphery than high-spatial frequencies (Pointer & Hess, 1989). Additionally, fixation durations should be shorter for high-spatial frequency targets, since high-spatial frequency targets are detected easier if they fall into the fovea than low-spatial frequency targets (Schütt & Wichmann, 2017). Another reason to prolong fixation durations for low-spatial frequency targets is that low-spatial frequency targets can be perceived from further away, and the size of the window in which targets can be detected increases with longer stimulus presentation (Geisler & Chou, 1995).

Thus, we expected a search behavior with small saccade amplitudes and short fixation durations when participants search for high-spatial frequency targets and a search behavior with large saccade amplitudes and long fixation durations when participants search for low-spatial frequency targets.

4.2 Results

We analyzed eye movement data from our experiment for search accuracy, search speed, average fixation duration, average saccade amplitudes and effects of changes in saccadic direction. All variables were investigated separately for the different search targets. Bar plots (left side of Figures 2–5) represent results for each of the 6 targets and the results for all targets combined in Session 7, when target type was unknown prior to each trial. Line graphs in Figures 2–7 show comparisons between the three low-spatial frequency targets

Figure 4.1: Illustration of the task. Subjects were asked to search for one specific target for a block of 25 trials, each overlayed over natural scenes like this one. In this image all 6 targets are hidden twice as large and at higher contrast than in the experiment, to make them visible despite the small image. In the actual experiment only one target was hidden per image and the image was shown much larger. The bottom panels show the 6 targets we used. The frames around the targets mark which frequency category they belong to.

(Gaussian Blob and positive/negative Mexican hat, black line) and the three high-spatial frequency targets (vertical, horizontal bar and cross, red line; cf. Fig. 4.1, bottom panels) throughout the course of the 8 experimental sessions. Error bars in the graphs are the standard error of the mean. Significance signs refer to differences between low and high-spatial frequency targets (* $p < .05$, ** $p < .01$, *** $p < .01$). Solid lines below significance stars indicate significant differences between low and high-spatial frequency targets for a range of neighboring data points (see Fig. 3–6).

4.2.1 Task performance

Detection rate

Performance of our group of 10 participants (see Methods) is characterized by similar detection rates (Hits/Misses) for the different targets throughout the whole experiment (Fig. 4.2A). The lowest detection rates were observed for the positive Mexican hat and the high-spatial frequency cross (both 83%) and the highest rate for the negative Mexican hat (92%). The overall rate of false alarms was very low (3.44% of target absent trials). Over the course of the experiment (Fig. 4.2B), the detection rate for both low and high-spatial frequency targets increased. No clear difference between the groups of high-spatial

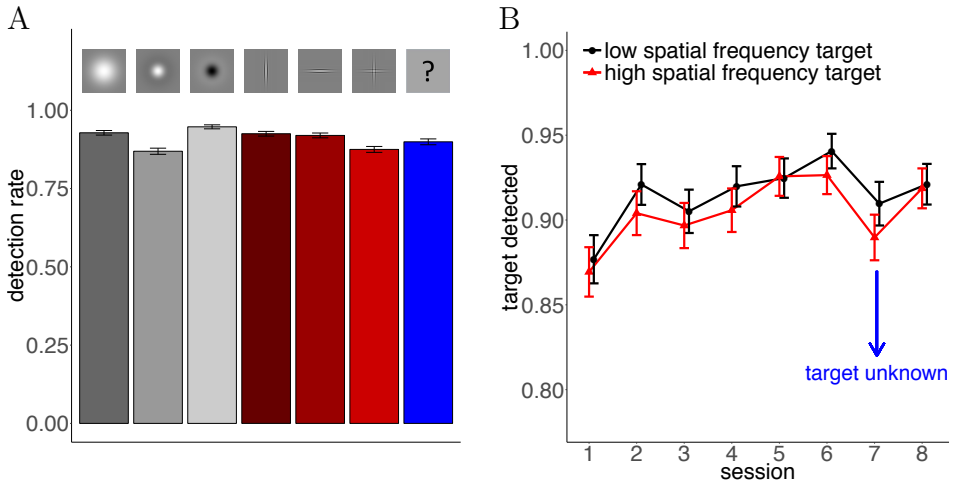

Figure 4.2: A) Detection rate for the 6 targets. Red bars are high-spatial frequency targets and gray bars low-spatial frequency targets. The blue bar captures all trials where target type was unknown prior to the trial. B) Average detection rate of the three low and high-spatial frequency targets throughout the 8 experimental sessions. In Session 7 target type was unknown prior to a trial.

frequency and low-spatial frequency targets was observed. In Session 7, when target type was unknown prior to the trial, detection rates dropped for both target types but performance was still better than in the first experimental session.

Search time

Mean search times (Fig. 4.3A) were more variable between the targets than the detection rate. Participants were faster at finding low-spatial frequency targets than high-spatial frequency targets. Participants were fastest at finding the negative Mexican hat and slowest at finding the high-spatial frequency cross. Search time decreased over the 8 sessions. The first 3 sessions showed a clear training effect and afterwards a plateau was reached (Fig. 4.3B). In Session 7 (target unknown) search times increased but high-spatial frequency targets were still detected faster than in the first session, indicating that search training compensated for loss of guidance in this case, which was also visible in detection performance.

4.2.2 Scanpath properties

Saccade amplitudes

Analyses of the saccade amplitudes throughout our experimental sessions (Fig. 4.4) showed three clear results: (i) Mean amplitudes were greater for low than for high-spatial frequency targets, (ii) this difference was established in the first session, persisted throughout all other sessions, and (iii) vanished when target type was unknown prior to a trial. Search-

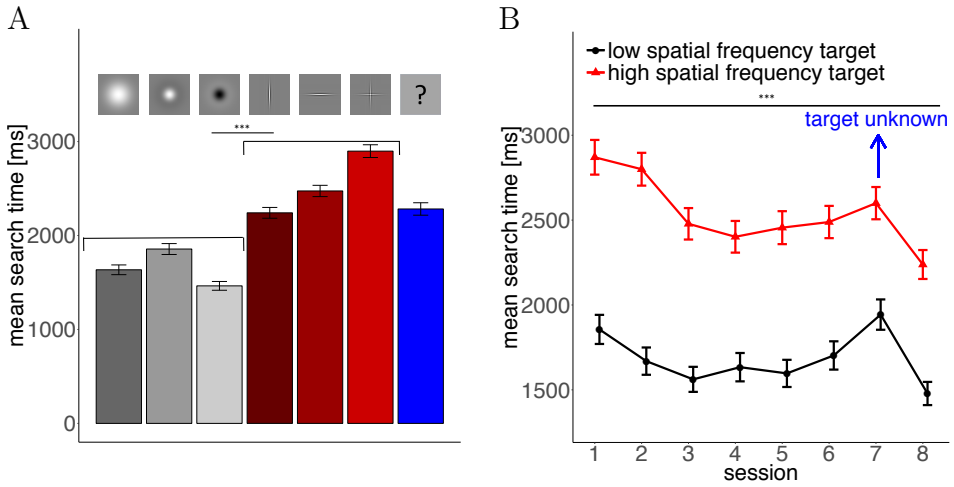

Figure 4.3: A) Search times for the 6 targets. Red bars are high-spatial frequency targets and gray bars low-spatial frequency targets. The blue bar captures all trials where target type was unknown prior to the trial. B) Average search times for the three low and high-spatial frequency targets throughout the 8 experimental sessions. In Session 7 target type was unknown prior to a trial.

ing for the Gaussian blob led to the largest mean saccade amplitudes (Fig. 4.4A). Overall, low-spatial frequencies produced larger saccade amplitudes, indicating that search strategy was adjusted to the visibility of the targets into the periphery.

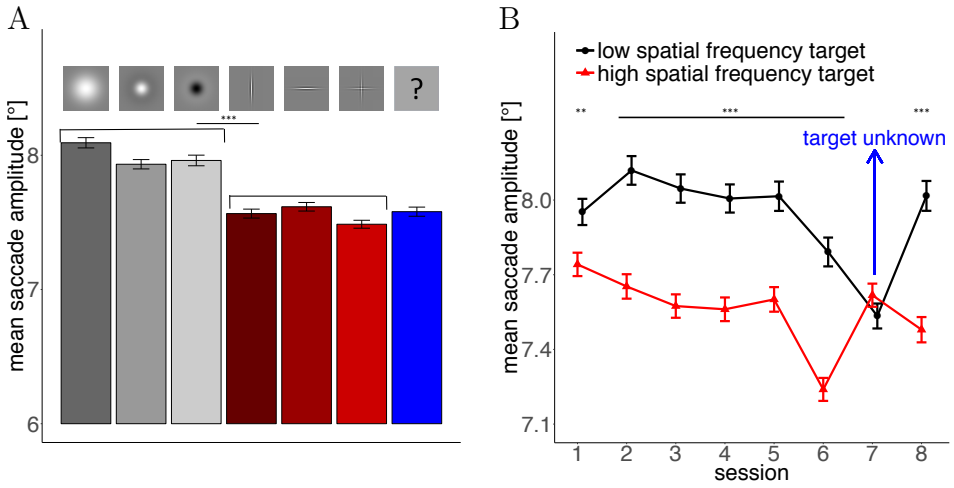

Figure 4.4: A) Mean saccade amplitude for the 6 different targets. Red bars are high-spatial frequency targets and gray bars low-spatial frequency targets. The blue bar captures all trials where target type was unknown prior to the trial. B) Average saccade amplitude of the three low and high-spatial frequency targets throughout the 8 experimental sessions. In Session 7 target type was unknown prior to a trial.

Fixation durations

The pattern for mean fixation durations (Fig. 4.5) was similar to the pattern of saccade amplitudes: (i) The three low-spatial frequency targets led to a search strategy with

longer fixation durations, (ii) this difference in fixation durations needed one training session to be established, but afterwards persisted throughout the other sessions, and (iii) vanished when target type was unknown prior to a trial. Fixation durations decreased throughout the experiment, thus mean fixation durations were rather short in Session 7, when the target was unknown prior to each trial (Fig. 4.5B). Again, the search strategy was adjusted according to the spatial frequency of the targets in a useful way, since it takes longer for low-spatial frequency targets to be detected when they fall into the fovea.

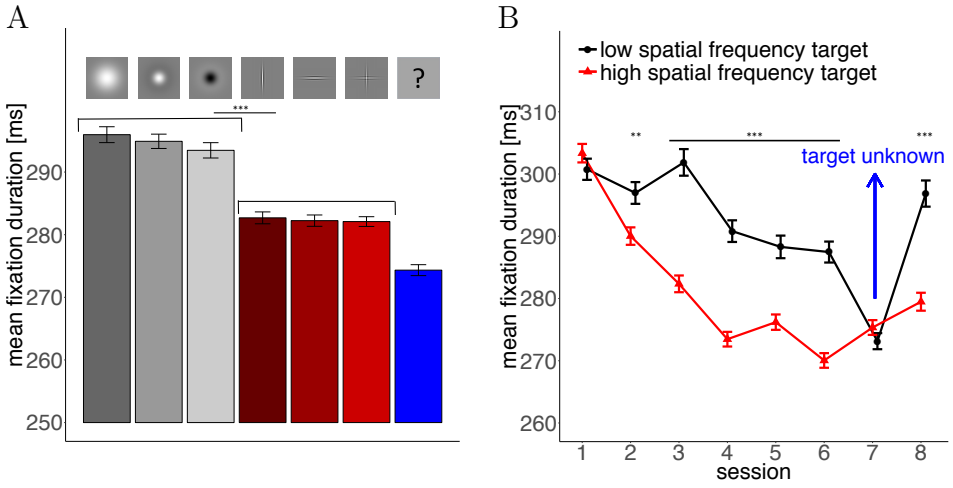

Figure 4.5: A) Mean fixation durations for the 6 different targets. Red bars are high-spatial frequency targets and gray bars low-spatial frequency targets. The blue bar captures all trials where target type was unknown prior to the trial. B) Average fixation duration of the three low and high-spatial frequency targets throughout the 8 experimental sessions. In Session 7 target type was unknown prior to a trial.

Time-course during a trial

In the sequence within a trial, mean fixation durations increased and mean saccade lengths decreased (except for the first fixations/saccades, which were influenced by the experimental design and the central fixation bias; Fig. 4.6). This behavior is known as the coarse-to-fine strategy of eye movements (Antes, 1974; Over et al., 2007). However, the effect of target spatial frequency already occured after the second saccade and lasted for the rest of the trial. Thus, participants displayed different coarse-to-fine strategies for low and high-spatial frequency targets.

Change in saccadic direction

Although we did not have a hypothesis about the interaction of saccade direction and visual search target, we included corresponding post-hoc analyses, since angles between successive saccades in scene viewing follow a very characteristic distribution (Tatler & Vincent, 2009; Smith & Henderson, 2009; Rothkegel et al., 2016) and strongly affect

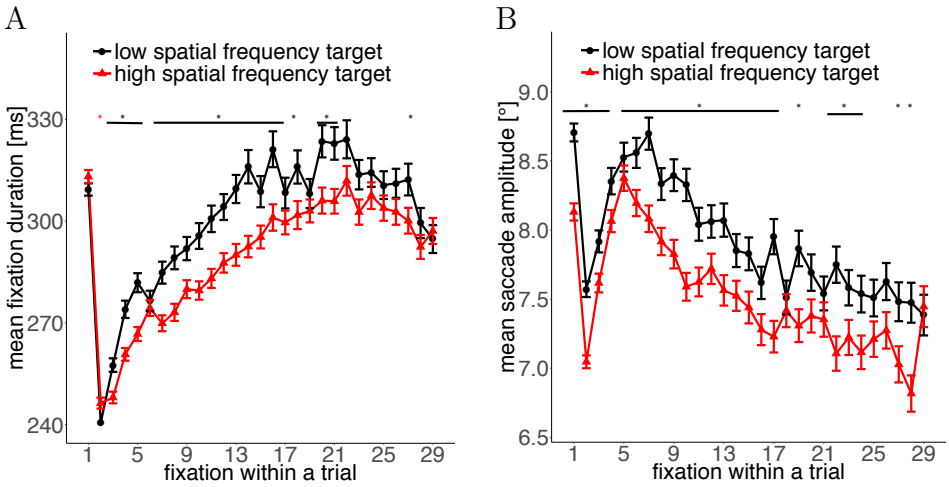

Figure 4.6: Temporal evolution of A) mean saccade amplitude and B) fixation duration for the two targets types throughout a trial.

saccade amplitudes and fixation durations (Tatler & Vincent, 2008; Tatler et al., 2017). Saccades which maintain direction (which we will denote as *saccadic momentum* saccades in the following (Smith & Henderson, 2009)) typically have small amplitudes and preceding fixation durations are short. Saccades with a 180 degree change in direction (or *return saccades*) are usually large and fixations, which preceed a return saccade, last rather long.

No apparent difference was found between the distributions of intersaccadic angles for low and high-spatial frequency targets (Fig. 4.7A). As has been described previously (Tatler & Vincent, 2008) we observered an increase in saccade amplitude as a function of change in saccade direction. Figure 4.7B also shows that saccades which maintained direction from the previous saccade (0 degree change) were equally large for both target types. Saccades which did contain a change direction differed in saccadic amplitude except for complete turns in direction (180 degree change). An influence of the angle is even more evident if we look at fixation duration differences between the two target types in relation to the change in saccade direction (Fig. 4.7C). Again, fixation durations increased for large changes in saccade direction. However, compared to previous results (Tatler & Vincent, 2008) the increase in our experiment was not linearly but reached a plateau for large changes in saccade direction an even decreased slightly for compelte return saccades. Again we see that saccades which maintain direction (0 degree change) show no difference for fixation durations between the two target types. For saccades which change direction, the fixation durations differ between the two target types. The fact that both, saccade amplitude and fixation duration do not differ for saccades without a change in direction led us to the hypothesis, that these saccades are less selective than other saccades and follow a sort of default scan.

To further investigate this hypothesis, we compared the saccadic landing points for different changes in saccade direction in terms of empirical density and visual saliency.

The empirical density maps were computed with the SpatStat package of the R language for statistical computing (Baddeley & Turner, n.d.; R Core Team, 2014) and for visual saliency we used the DeepGaze 2 model (Kümmerer et al., 2016), which is the currently highest ranking saliency model on the MIT saliency benchmark (Bylinskii et al., 2015). The fixations were evaluated in terms of their likelihood under the Deep Gaze model or the empirical density compared to a uniform distribution (see (Kümmerer et al., 2015; Schütt et al., 2017) for further elaboration). Values above zero thus indicate improvements compared to a uniform distribution, negative values represent predictions below chance.

Figures 4.7D and E show that saliency and empirical density, respectively, depended on the previous change in saccade direction. Saccade targets were most salient (Fig. 4.7D) and visited more by all other participants (Fig. 4.7E) if the previous saccade had a large change in direction (180 degree). Saliency values increased continuously with larger changes in saccade direction. Empirical density of the saccade targets was highest for return saccades (180 degree) but lowest for saccades with a left or right turn (90 degree) from to the previous saccade. This did not match our hypothesis, that the least selective saccades are saccades which maintain direction (0 degree change). However, most saccades wich maintain direction are rather short, and short saccades often land at highly interesting positions, because they contain corrective saccades. Thus, we analyzed the empirical density with respect to the preceding change in direction for different saccade amplitudes separately. Figure 4.7F shows the the empirical density with respect to previous change in saccade direction only for saccades between 3 and 8 degrees of visual angle. Removing the rare large saccades and small corrective saccades led to the smallest empirical density for saccades without a change in direction, as hypothesized for a default scan mechanism. If we conduct this analysis for one degree bins of saccade amplitude sizes separately, the increase of empirical density for an increasing change in saccade direction is evident for all amplitudes between 2 and 11 degree. Larger saccades show a rather noisy distribution and smaller saccades a rather constant value, independent of the previous change in saccade direction.

Although visual saliency depended on the change in direction (Fig. 4.7D), all fixations are predicted below chance by the visual saliency model (the difference in log-likelihood compared to a uniform prediction was negative). This agrees with the notion that visual saliency does not predict fixation locations in visual search above chance (Henderson et al., 2007; Schütt et al., 2018).

4.2.3 Spatial frequency spectra of fixated locations

Earlier analyses of eye movements during visual search reported similarities between the fixated locations and the target and, consequently it was assumed that such relationships

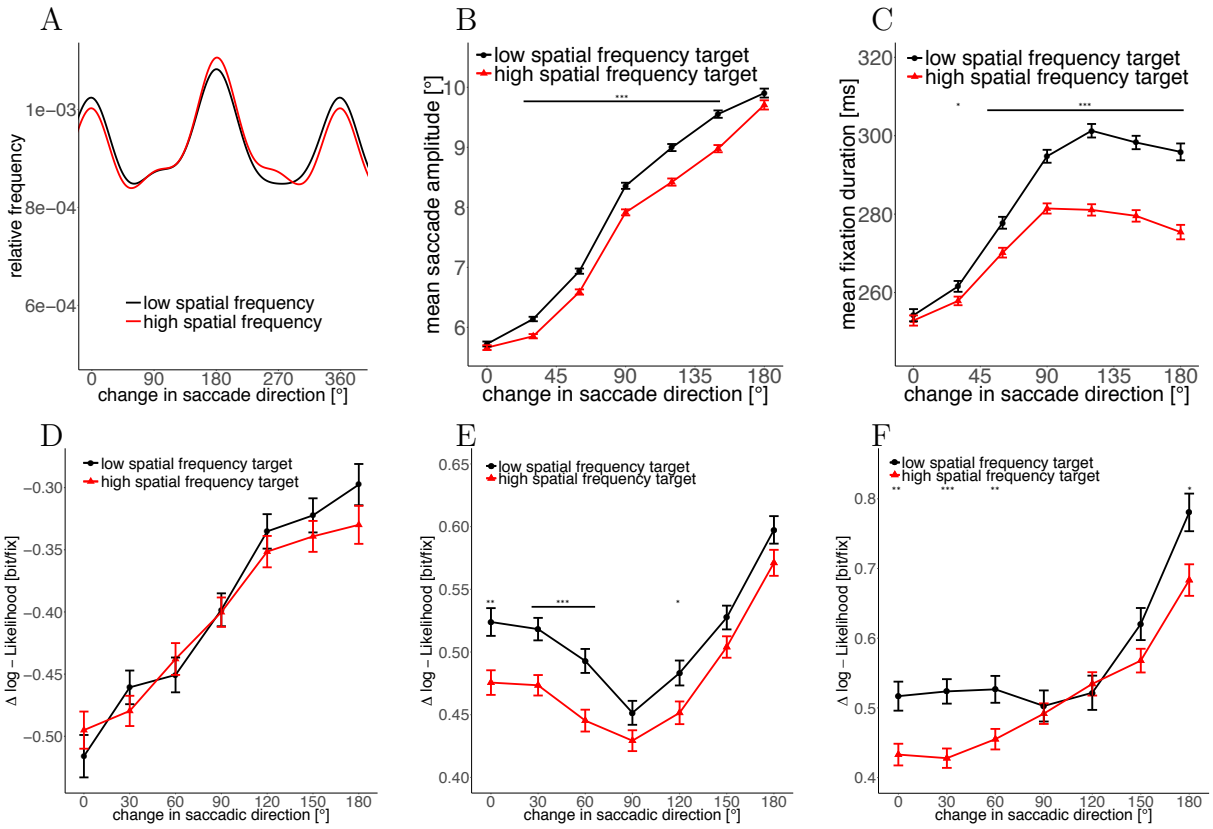

Figure 4.7: A) Density distribution of change in saccadic direction. Influence of change in saccadic direction on B) successive saccade amplitude C) fixation duration between the two saccades D) DeepGaze saliency of successive saccade target E) empirical density of successive saccade target and F) empirical density of successive saccade target only for saccades between 3 and 8 degree of visual angle.

could be exploited for the prediction of fixation locations (Hwang et al., 2009). Thus, we investigated whether a corresponding difference between fixated and non-fixated image locations exists in our data. As fixated locations, we extracted patches around the fixation locations and compared them to control patches extracted from the same locations in a different image from the stimulus set (Judd et al., 2009; Kienzle, Franz, Schölkopf, & Wichmann, 2009) (see Methods). To compare the fixated patches for the different targets, we analyzed the spectra of the patches (Fig. 4.8; see Methods). As displayed in Figure 4.8A, the average spectrum of a fixated patch looks much like the spectrum of any image patch with a clear $\frac{1}{f}$ decline in spatial frequency content and a preference for horizontal and vertical structure. As these strong effects hide all other effects, all other spectra are divided by the spectra of the comparison patches for display.

The overall spectrum of fixated patches shows increased power for all frequencies and orientations (Fig. 4.8B) compared to a random image patch, indicating that fixated patches have more contrast than non-fixated patches. The unknown target condition (Fig. 4.8C) produces no clear deviation from the average over the conditions with known target. Searching for a specific target produces a slight bias of the fixated image patches towards being more similar to the spectrum of the target (Fig.4.8 D). The deviations of the single targets from the grand average are all smaller than 5%, however, while the variance over patches is substantial ($\frac{SD}{M} \in [78.65\%, 161.03\%]$, average $= 91.10\%$).

While these results indicate a bias towards image patches, which have similar spectrum to the target, differences in the range of 0.1 standard deviations are certainly too small to infer the fixation category from the spectrum. Thus the only distinction, which might provide some predictive value is the generally increased contrast at fixated locations in general.

4.2.4 Target difficulty

Since we placed the targets on different, pseudorandom positions in the images, it was - by chance - sometimes easy and sometimes hard to find them. A direct measure of how difficult it was to find a target, is the time it took participants to find the target. As a computational measure, we used a recently published early vision model for images, which computes a signal-to-noise-ratio (SNR) of the target on the background for all target patches (Schütt & Wichmann, 2017), (i) to evaluate whether the model can predict search behavior on natural scenes and (ii) to obtain a measure of visibility for each of the targets. A glance at Figure 4.9 indicates that both, detectability and search time were correlated with the predicted SNR from the early vision model. The computed SNR by the early vision model thus predicts search behavior. The model computes SNRs for foveal vision. Note however, that the high-spatial frequency targets have higher SNRs than the

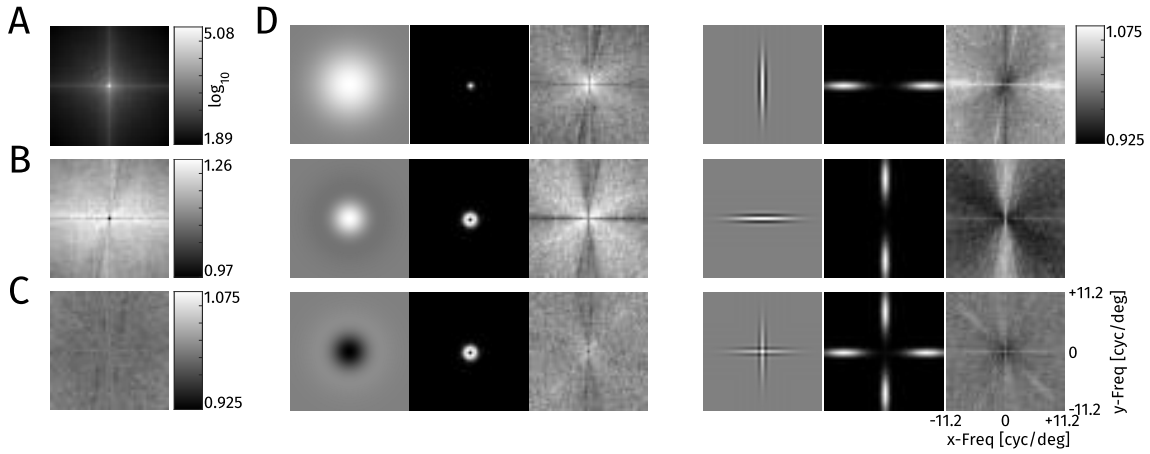

Figure 4.8: Analysis of spatial frequency amplitude spectra at fixated locations. **A**: Grand average spectrum over all fixated patches. **B**: Spectrum from A divided by the average spectrum at control locations. The value at 0 frequency is 0.97, all other values are in the range of $[1.09, 1.26]$ **C**: Average spectrum for fixations when the target is unknown, plotted as for known targets in D. **D**: Triples for each target: The target at 100% contrast against a gray background, the amplitude spectrum of the target and the average amplitude spectrum at fixation locations divided by the average over all targets. The color range from black to white for the third plot is always $[0.925, 1.075]$.

low-spatial frequency targets and are thus easier to see, when looked at in foveal vision (different scales of y-axis in Fig. 4.9). Nonetheless, low-spatial frequency targets were found significantly faster than high-spatial frequency target, arguing that the periphery and eye movements play a highly important role in visual search (Nuthmann, 2014).

4.3 Discussion

We studied visual search for artificial low and high-spatial frequency targets in natural scenes and found that fixation durations and saccade amplitudes depend on the low-level properties of the search target. The different influences of a target on these basic eye-movement characteristics are part of a top-down search strategy, since differences between targets disappeared immediately, as soon as participants did not know which target to search for. Additionally, differences between target types also occurred when the target was absent but participants were told which target to look for. Our findings imply that humans adjust their basic search behavior to the target they look for. In our study, fixation durations and saccade amplitudes were longer for low-spatial frequency targets. Previous research has shown that detectability of targets in the periphery depends on spatial frequency (Pointer & Hess, 1989) and fixation duration (Geisler & Chou, 1995). Increasing fixation durations thus lead to a larger window of detectability and low-spatial frequency targets can generally be detected from further away. For high-spatial frequency

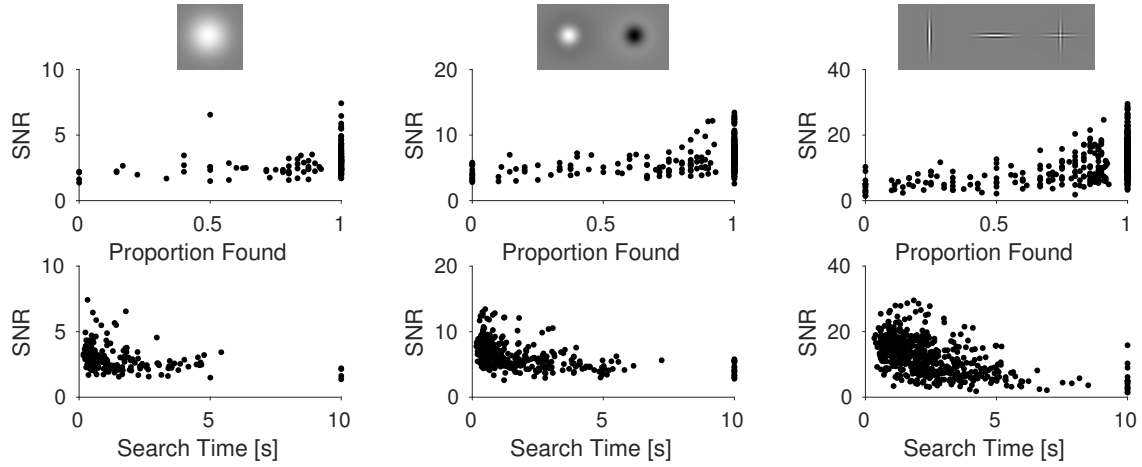

Figure 4.9: Signal to Noise Ratio from an early vision model for all target-background combinations (Schütt & Wichmann, 2017). Search times and detectability are correlated, when separating analysis between target types.

targets, participants decreased their fixation durations strongly over time. Decreasing fixation durations and thus increasing frequency of fixations for high-spatial frequency targets is a useful search strategy because they (i) cannot be detected from far away and (ii) have a higher SNR when looked at directly (see Fig. 4.9).

A recently published model of early vision (Schütt & Wichmann, 2017) was used to predict target difficulty. Separately looking at the results for the different targets produced promising results. We found high correlations between a targets signal-to-noise-ratio and search times and detectability values. However, the targets themselves had very different signal-to-noise-ratios, which were not reflected in search times. One possible reason that the model fails at predicting search times between different target types is, that only foveal input is modeled. The fact that low-spatial frequency targets are found faster than high-spatial frequency targets, although they have a lower SNR, shows that the periphery plays a strong role when searching for artificial targets on complex backgrounds.

Analyses of the fixation locations demonstrate that searchers slightly adjust where they look to depending on the target and confirming earlier reports (Wolfe, 1994; Hwang et al., 2009). However, the influence of the target on fixation locations (investigated by comparing spatial frequency spectra) is rather small, agreeing with the notion that participants do not merely look at positions which mostly resemble the target (Najemnik & Geisler, 2005), but take their peripheral vision into account.

Additional post-hoc analyses of the changes in saccade direction and its dependence on further viewing behavior revealed interesting results. Saccades, which generated strong changes with respect to previous scanpath direction, landed at locations with higher empirical fixation probability (when removing small corrective saccades from the analysis) and visual saliency, and were influenced by target properties, except for saccade ampli-

tudes after a 180 degree change in direction. Saccades which maintained direction were not influenced by the search target and corresponding endpoints had low fixation density and low saliency values. For the control of fixation durations, our current results lend support to the concept of mixed control (Henderson & Pierce, 2008; Trukenbrod & Engbert, 2014), meaning that the visual input as well as some independent time-keeper influences when a saccade is generated. The eyes simply progress in a default manner with *saccadic momentum* saccades, unless an interesting object captures the attention, which prolongs fixation duration and urges the eyes to change direction. The idea, that interesting locations can inhibit default *saccadic momentum* saccades, strongly suggests that eye movement behavior in search is shaped not only by foveal, but also peripheral vision (Nuthmann, 2014; Cajar, Engbert, & Laubrock, 2016; Hulleman & Olivers, 2015).

We interpret our results as an indication for two different ways of searching, a selective search and a default scan, which primarily moves the eyes forward in one direction. These findings agree with a study by Bays and Husain (Bays & Husain, 2012), who reported that return saccades are generally inhibited and only executed if a saccade target is highly interesting while forward saccades are facilitated and more frequent than a random, memoryless control mechanism would predict.

Eye movements play a substantial role for visual search in natural scenes and are at least partially under top-town control. However, there also seems to be a default scanning mechanism, which continues to move the eyes in the previous saccade direction and is not adjusted to needs of the target-template. This default scan might simply be the result of an evolutionary program to facilitate foraging (Wilming et al., 2013). Thus, our results are consistent with at least two mechanisms controlling eye movements under natural search conditions, which is important for dynamical models of scanpath generation (Engbert et al., 2015; Le Meur & Liu, 2015; Tatler et al., 2017).

4.4 Methods

We generated 6 different low-level targets with different orientation and spatial frequency content (Fig. 4.1):

A *Gaussian blob* with a standard deviation of 0.4° of visual angle. This is an isotropic stimulus, which is a Gaussian in spatial frequency as well (with a standard deviation of $\sigma_f = 0.3979$). A *positive Mexican hat*, the difference between a Gaussian with a standard deviation of 0.2° and a Gaussian with standard deviation 0.4°. This stimulus is isotropic and has a peak frequency of roughly $0.7\frac{cyc}{deg}$. A *negative Mexican hat*, the negative of the positive Mexican hat, which has exactly the same spatial frequency spectrum. A *vertical Gabor*, the product of $8\frac{cyc}{deg}$ vertical cosine centered at the origin and a Gaussian with standard deviations of 0.06° and 0.32° in x and y direction. In frequency space

this stimulus is strongly oriented and has a relatively broad frequency peak at $8\frac{cyc}{deg}$. A *horizontal Gabor*, the same as the vertical Gabor but oriented horizontally. A *Gabor cross*, the sum of the two Gabors, each at half the contrast.

All stimuli but the Gaussian blob were near zero mean and all stimuli were normalized to have an amplitude of 1, i.e. $max(abs(T)) = 1$.

To embed the targets into the natural images, we first converted the image to luminance values based on a power function, fitted to the measured luminance response of the monitor. We then combined this luminance image I_L with the target T with a luminance amplitude αL_{max}, fixed relative to the maximum luminance displayable on the monitor L_{max} as follows:

$$I_{fin} = \alpha L_{max} + (1 - 2\alpha) I_L + \alpha L_{max} T. \tag{4.1}$$

We rescaled the image to the range $[\alpha, (1 - \alpha)]L_{max}$ and then added the target with a luminance amplitude of αL_{max}, such that the final image I_{fin} never left the displayable range. We then converted the image I_{fin} back to $[0, 255]$ grayscale values by inverting the fitted power function.

4.4.1 Target locations

For placement of the targets we lay a grid of 4×2 rectangles over each image. Within each rectangle, we chose a random position for each target and image, which was at least 100 pixels away from the border, such that the target was not cut off at any side. The original plan was to present each target at each position in each image once over the eight sessions of one observer. Unfortunately, a bug in the experimental code led to a random choice of the target location instead, but we sampled only among the 8 possible locations sampled for the target-image combination. Most target-position-image combinations appeared between 6 and 10 times (10 participants and about 20 % target absent trials, mean=7.8) and none was present more than 16 times. We are rather certain that participants could not remember the position-target-image combinations over 1200 trials and even if they did, a target appeared so rarely again at the same position that it would not have been strongly predictive of the target position. Furthermore, no participant mentioned noticing anything like repeating target positions.

4.4.2 Experiment

Stimuli

As stimulus material we used 25 images taken by L.R. and a member of the Potsdam lab with a Canon EOS 50D digital camera (max. 4752 x 3168 pixels). The images were outdoor scenes without people, animals or written words present. Most images had parts

with a lot of high-spatial frequency content (grass or woods) and parts with no high-spatial frequency content (sky or empty street). They were all taken on a bright sunny day in the summer.

Stimulus Presentation

Stimuli were presented on a 20-inch CRT monitor (Mitsubishi Diamond Pro 2070; frame rate 120 HZ, resolution 1280×1024 pixels; Mitsubishi Electric Corporation, Tokyo, Japan). All pictures were reduced to a size of 1200×960 pixels. For the presentation during the experiment, images were displayed in the center of the screen with gray borders extending 32 pixels to the top/bottom and 40 pixels to the left/right of the image. Images covered 31.1 degree of visual angle in the horizontal and 24.9 degree in the vertical dimension.

Participants

We recorded eye movements from 10 human participants (4 female) with normal or corrected-to-normal vision in 8 separate sessions on different days. 6 participants were students from a nearby high school (age 17 to 18) and 4 were students at the University of Potsdam (age 22 to 26). The work was carried out in accordance with the Declaration of Helsinki. Informed consent was obtained for experimentation by all participants.

Procedure

Participants were instructed to position their heads on a chin rest in front of a computer screen at a viewing distance of 70 cm. Eye movements were recorded binocularly using an desktop mounted Eyelink 1000 video-based-eyetracker (SR-Research, Osgoode/ON, Canada) with a sampling rate of 1000 Hz. Participants were instructed to search a target for the upcoming 25 images. Before each block of 25 images, the target was presented on an example image, marked by a red square. Each session consisted of 6 blocks with 25 images with the 6 different targets. The 25 images were always the same images.

Overall, 10 participants searched 6 targets on 25 images in 8 sessions, thus we collected data of 12000 search trials. Target absent trials made up between 3 and 7 for each block of 25 images ($\sim 80\%$).

Trials began with a black fixation cross presented on gray background at a random position within the image borders. After successful fixation, the image was presented with the fixation cross still present for 125 ms. This was done to assure a prolonged first fixation to reduce the central fixation tendency of the initial saccadic response (Tatler, 2007; Rothkegel, Trukenbrod, Schütt, Wichmann, & Engbert, 2017). After removal of the fixation cross, participants were allowed to search the image for the previously defined

target for 10 s. Participants were instructed to press the space bar to end the trial, once a target was found.

At the end of each session participants could earn a bonus of up to 5€ additional to a fixed 10€ reimbursement, depending on the number of points collected. Participants earned 1 point for each correctly identified target. If participants pressed the bar although no target was present, one point was subtracted.

Data preprocessing and saccade detection

For saccade detection we applied a velocity-based algorithm (Engbert & Kliegl, 2003a; Engbert & Mergenthaler, 2006). This algorithm marks an event as a saccade if it has a minimum amplitude of 0.5 degree and exceeds the average velocity during a trial by 6 median-based standard deviations for at least 6 data samples (6 ms). The epoch between two subsequent saccades is defined as a fixation. All fixations with a duration of less than 50 ms were removed from further analysis since these are largely glissades (Nyström & Holmqvist, 2010). The number of fixations for further analyses was 166,903.

4.4.3 Fixation locations analysis

Empirical density and saliency at saccadic endpoints

To estimate empirical fixation densities, we used kernel density estimation as implemented in the R package SpatStat (version 1.51-0). To estimate the bandwidth for the kernel density estimate we used leave one subject out cross-validation, i.e. for each subject we evaluated the likelihood of their data under a kernel density estimate based on the data from all other subjects repeating this procedure with bandwidths ranging from .5 to 2 degrees of visual angle (dva) in steps of 0.1 dva. We report the results with the best bandwidth chosen for each image separately. We then took the resulting density value of each saccade target and averaged for the different target types and previous changes in saccade direction. Likelihood values are the average of each fixation position on the map, taken from a grid of 128×128 grid cells. The DeepGaze II model (Kümmerer et al., 2016) provides a map, where we could simply draw saliency values for each fixation. We again averaged these values for the different target types and previous changes in saccade direction.

For further information on likelihood evaluation of saliency models we refer to Kümmerer et al. (2015), Schütt et al. (2017) and Schütt et al. (2018).

Spatial frequency spectra

To analyze the image properties at fixation locations, we extracted image patches around fixation locations and compared them over targets and to comparison locations. We extracted 79×79 pixel patches ($\approx 2.05 \times 2.05$ dva), around the fixated pixel, for all fixation locations for which this patch lay entirely inside the image. To obtain comparison patches, we extracted patches at the measured fixations locations shifting the image index by one, i.e., we used the fixations from picture one to extract patches from picture two (and so on), and the fixations from the last picture to extract patches from the first picture, as was done earlier to train saliency models (Judd et al., 2009; Kienzle et al., 2009).

For our analysis, we converted the patches to luminance using the measured gamma curves of the screen and calculated the spatial frequency spectrum using MATLAB's fft2 function. Then we calculated the amplitude as the absolute value for each frequency and averaged it over patches within a group to display. To display differences between conditions, we divided the average of one group by the average of the other. To quantify the variability of patches within one condition we divided the standard deviation of amplitudes by the mean value ($\frac{SD}{M}$).

4.5 Acknowledgements

This work was supported by Deutsche Forschungsgemeinschaft (grants EN 471/13-1 and WI 2103/4-1 to R. E. and F. A. W., resp., and CRC 1294). We thank the members of the EyeLab at the university of Potsdam for conducting the experiment and collecting the data. We thank Anke Cajar for reviewing the manuscript prior to submission and Daniel Backhaus for assisting with the creation of the stimulus material.

4.6 Author contributions statement

L.O.M.R and H.H.S programmed the experiment, analyzed the data and wrote the manuscript. All authors developed the idea for the experiment. All authors reviewed the manuscript.

4.7 Competing financial interests:

The authors declare no competing financial interests.

Chapter 5

General discussion

Eye-movement data from laboratory scene-viewing experiments are investigated to increase our understanding of active vision in the real world. One primary goal of scene-viewing research is to determine where and how visual attention in a scene is allocated. Many years of investigating where observers look at in an image, have led to computational models which can explain a large amount of variance in scene-viewing data (Itti et al., 1998; Borji & Itti, 2013; Kümmerer et al., 2015). The question of how long participants look at certain locations has also gotten into the focus of research and in the last decade, elaborate models for fixation durations have been proposed (Nuthmann et al., 2010; Laubrock et al., 2013). However, even by knowing exactly where participants look at and how long they look there, the order of fixations within one scanpath cannot be predicted adequately (Rothkegel et al., 2016; Schütt et al., 2017). Dynamical models, which take fixation history of a scanpath and the inhomogeneity of acuity within the visual field into account, have been developed to overcome this obstacle (Engbert et al., 2015).

The purpose of the present thesis was to obtain new information about systematic eye-movement behavior, some of which can only be found when investigating dynamic aspects of scanpaths from scene-viewing experiments. By systematically investigating three aspects of natural scene perception, the role of initial fixation position, the temporal evolution of the central fixation tendency, and the influence of low-level target features on basic scanpath properties in complex visual search, we were able to answer several questions to improve and evaluate our dynamical model of scanpath generation in scene viewing, the SceneWalk model.

5.1 How do results fit into the existing literature?

Before providing an outlook for what the results from Chapters 2–4 imply for future research of scene viewing, I will summarize the results and embed them into current viewpoints.

5.1.1 Inhibition of return in scene viewing

In Chapter 2 we approached one of the key questions regarding dynamic aspects of human scanpaths: What is the force, which urges the eyes to move through an image and prevents them from oscillating between the most interesting (or salient) points on an image? Dynamical models often implement this force in the form of spatial inhibition of return (IOR). Spatial IOR is an attentional mechanism which inhibits the eyes from refixating previously examined locations (Posner et al., 1985; Klein, 2000). However, there has been a strong debate about whether spatial IOR exists in natural scene viewing.

A large amount of perfect return saccades back to the last fixation position (Smith & Henderson, 2009; Rothkegel et al., 2016), found in data from scene-viewing experiments, favored the idea that IOR has no influence on the selection of fixation locations. Smith and Henderson (2009) showed that a sudden luminance onset at the previous fixation position attracted more saccades, than if it appeared at positions, which would induce a left or right 90° turn of the saccade. They considered this proof that the previous fixation is not inhibited but even facilitated and thus termed their finding *facilitation of return*. To confirm their theory, they showed that empirical scene-viewing data contain more return saccades than surrogate data, which takes human saccade amplitudes and angles into account. However, Bays and Husain (2012) showed that, when taking saliency values and typical angle distributions of saccades into account, return saccades are in fact inhibited and not facilitated. They additionally showed that saccades which maintain direction are facilitated strongly. This facilitation of saccades in the same direction as preceding saccades has been termed *saccadic momentum* (also see Smith & Henderson, 2009; Wilming et al., 2013; Luke et al., 2014). In Chapter 2 we have shown that the initial fixation with an experimentally prolonged duration was inhibited on later fixations. This was shown by a strong overshoot to the image side opposite to the starting position. This overshoot was not found in simulated data from many statistical models without an inhibition of return mechanism and, more importantly, also not found in a model incorporating a saccadic momentum mechanism. The SceneWalk model, which uses the fixation map as an inhibition of return mechanism, was the only model which produced the characteristic overshoot observed in experimental data. Additionally, a facilitation of return back to the starting position with a long fixation was not observed in the data. Smith and Henderson proposed that inhibition of return only influences the last or second

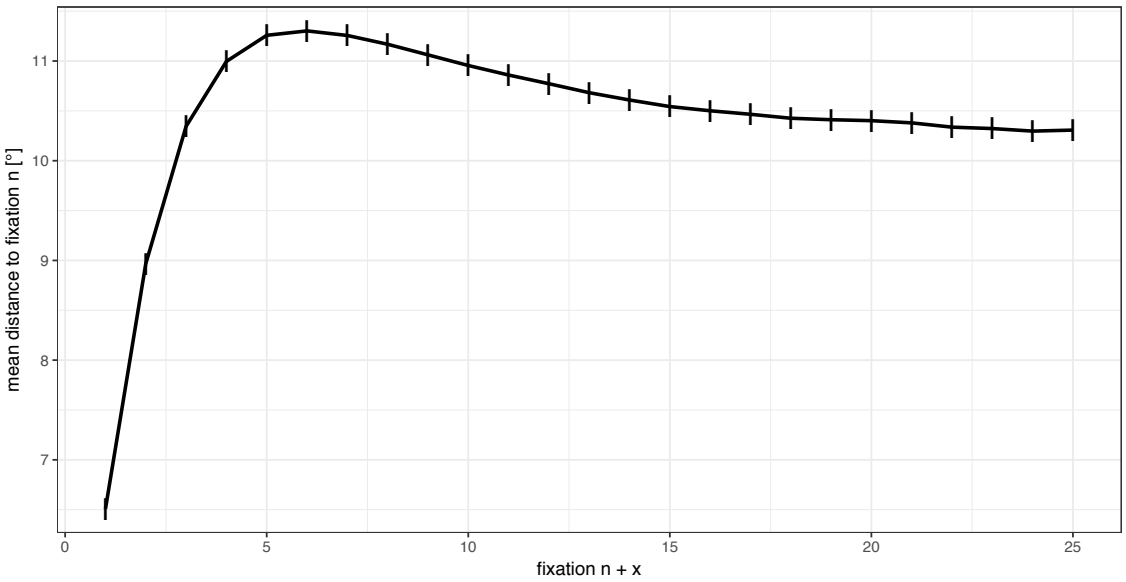

Figure 5.1: Mean distance from one fixation to successive fixations. Data is taken from a large Corpus study from our lab with 90 images and 105 observers (Schütt et al., 2018).

last fixation. The overshoot from our experiment is active for up to five seconds. We thus argue that inhibition of return in scene viewing is a more long-lasting effect which evolves later than in highly controlled saccade tasks. In a large corpus study (see Schütt et al., 2018) we found that the average distance of successive fixation positions increases for the six upcoming fixations and then decreases until a stationary value is reached (see Fig. 5.1), which also indicates that inhibition of return has a long-lasting influence on eye movements in natural scene viewing. Thus, inhibition of return, as used in the SceneWalk model, appears to be a viable mechanism to guide the eyes through an image.

5.1.2 The influence of the angle between two successive saccades

In Chapters 2 and 4 we have replicated results from previous studies, that the distribution of angles between successive saccades has a very rather characteristic shape (e.g., Smith & Henderson, 2009; Tatler & Vincent, 2009). In Chapter 4 we have shown that fixations, which follow saccades with a large change in direction, are visually more salient and fixated more often than other locations. This indicates that a change in saccade direction is only conducted when the target location is of special interest. The opposite was found for

landing positions of saccades, which maintain their current direction (*saccadic momentum saccades*). They are on average less informative and are not influenced by the low-level features of search targets. Thus, I believe that additionally to spatial IOR, a default scan which mostly contains saccades in the same direction exists in scene viewing. This scan can be explained as a mechanism which facilitates foraging (Wilming et al., 2013). The default saccadic momentum scan follows a sort of reading-like viewing behavior, which is interrupted, when interesting saccade targets prevent the next default saccade from happening. An open question that remains unanswered is why so many saccades show a complete turn in direction, compared to other saccadic angles. We have shown in Chapter 2 that all models produce a large amount of saccades with a complete turn in direction. I think that this is a result of the image framing, forcing the eyes to stay within certain boundaries. Additionally, it is physiologically not possible to constantly move the eyes in one direction. However, it is possible to change the direction after each saccade. Thus, the peak of backward saccades in the distribution of angles between successive saccades might simply be an oculomotor, experimentally enhanced necessity, whereas the peak of saccades which maintain direction is not.

5.1.3 The central fixation bias in scene viewing

One of the most prominent systematic eye-movement tendencies in natural scene viewing is the central fixation bias (Buswell, 1935; Tatler, 2007). This strong tendency is in fact the single best feature to predict fixation locations in scene-viewing experiments (Vincent et al., 2009). We were interested in the main cause of this bias, because it spans across all images and observers. We found that the initial saccade latency of an observation can predict the central fixation bias on a trial by trial basis. Thus, we argue that the artificial situation in the laboratory, with a sudden image onset, is one of the main contributors to this bias. We are aware that a photographers bias (Tatler, 2007), the head position (Vitu et al., 2004), the framing of the image on the screen (Bindemann, 2010), and other factors also contribute to this bias. However, all explanations agree with the notion that the center bias is a laboratory artifact. In mobile eye-tracking, the central fixation bias is reduced ('t Hart et al., 2009). However, in mobile eye-tracking studies, center bias is equivalent to the eyes looking at the center of the visual field (i.e. the camera of the eye tracking device. See Ioannidou et al. 2016), which is not the same as looking at the center of an external object, e.g., the image on the screen. Thus, the question, whether an object-based central fixation bias exists outside of the laboratory situation, has not yet been answered to my knowledge.

The results presented in Chapter 3, indicate that the central fixation bias hides or overrides attentional processes elicited by the image or the task. This central fixation

bias can be reduced by dissociating the sudden image onset from initial eye movements.

5.1.4 Adaptiveness of the visual system to search targets

In Chapter 4 we tested whether eye movement parameters adjust to the visual properties of a target in visual search. Although previous work has shown that properties of the search template (i.e. the target) and fixated locations correlate with respect to visual features (Hwang et al., 2009), no studies have yet investigated how target features influence fixation durations and saccade amplitudes. We found that the spatial frequency of a target influences both saccade amplitude and fixation durations, such that saccade amplitudes are larger and fixation durations are longer for low than for high-spatial frequency targets. Both adjustments confirm a tight coupling between perception and eye movements. The oculomotor system is adjusted in a way, which facilitates search in this experiment. Large saccade amplitudes are useful for low-spatial frequency targets, since low-spatial frequency targets are more visible in the periphery than high-spatial frequency targets. Fixation durations are also longer for low-spatial frequency targets, probably because in foveal vision high-spatial frequency targets can be detected better than low-spatial frequency targets (Schütt & Wichmann, 2017) but not in peripheral vision. The adjustment of the eye movement characteristics can also be seen in target absent trials, indicating that the adaption of gaze control is driven in a top-down manner. Additionally, in target present trials, when target identity was unknown to the searcher, differences between target types vanished. Overall, results from Chapter 4 show that basic eye-movement characteristics are adjusted to the visual properties of the search target in complex visual search.

5.2 Implications for future research

5.2.1 Implications for scene-viewing experiments

The results from this thesis lead to new ideas for the design of scene-viewing experiments.

First, the central fixation bias should be avoided as much as possible, because it overrides top-down and bottom-up influences on fixation selection. We reduced the bias by introducing a pre-trial fixation time, during which participants are forced to view the image from a predefined position for a variable duration, before they are allowed to explore the image. We recommend to use a pre-trial fixation time of 125-250 ms to reduce the influence of the sudden image onset. It is also possible to evaluate eye movements from scene viewing beyond the central fixation bias, by incorporating the bias as an image independent baseline (Clarke & Tatler, 2014; Nuthmann et al., 2017). The initial fixations of an observation are driven strongest by bottom-up features (Parkhurst et al.,

2002; Anderson, Donk, & Meeter, 2016; Schütt et al., 2018) and thus, reducing the bias on initial fixations can enhance our understanding of factors, which drive the first voluntary eye movement beyond the sudden image onset. Another possibility to reduce the bias is to use large, almost borderless pictures. This can be done with virtual reality devices or mobile eye tracking. Novel techniques will help to answer the question, whether an object-based central fixation bias exists in natural viewing behavior.

Besides the duration of the initial fixation, the location of the starting position of an observation influences further scanpath progression strongly. We showed in Chapter 2 that each observation depended on the starting position for up to 5 seconds. In most studies, the starting position is in the center of an image and afterwards neglected, when investigating the data. Although probably not the perfect solution for this problem, a variable starting position reduces systematic biases. Whatever starting position is chosen, it is crucial for further analysis that the impact of the initial fixation on further scanpath progression is known to the experimenter.

In visual search on complex scenes, target properties need to be evaluated before the experiment, because factors like visibility in the periphery will shape eye movements (Chapter 4) and thus one visual search path is not simply transferable to another one, when for example fixation durations are investigated.

The last aspect I want to mention here is the amount of viewing time for each trial. Throughout a trial viewing behavior changes systematically. We showed in Chapter 2 that the starting position influences further viewing behavior for up to 5 s. Additionally, results from Chapter 3 have shown that the first second of viewing is strongly influenced by the central fixation bias, especially when the starting position is in the center. Also, most experiments show a coarse to fine strategy, meaning that fixation durations get longer and saccade amplitudes get shorter with increasing viewing time (Over et al., 2007; Rothkegel, Schütt, Trukenbrod, Wichmann, & Engbert, 2018). We have also shown that a large between-subject congruency of fixation locations is only observed in the first 1-2 seconds and decreases systematically after the first fixation (Schütt et al., 2018). Thus, data from experiments with long-lasting observations have to be treated different than data from short observations. To understand the temporal dependencies, it is helpful to investigate all obtained results as a function of viewing time.

Unfortunately, even if all afore mentioned aspects of the experiment are controlled for and a scene-viewing experiment is conducted in a manner which tries to avoid all biases, it needs to be questioned how eye-movement behavior, when looking at a photo on a screen, can transfer to the behavior, when seeing the same scene in the natural environment (Tatler et al., 2011).

5.2.2 Implications for computational models

The most common approach when modeling visual attention and eye movement behavior in scene viewing has been to compute bottom-up feature maps, which predict fixation locations (Itti et al., 1998; Borji & Itti, 2013). Many researchers have pointed out that only computing bottom-up saliency is not sufficient to predict fixation locations in a scene, because higher-level features, like objects, outperform low-level visual saliency (Einhäuser et al., 2008; Nuthmann & Henderson, 2010; Kümmerer et al., 2016; Schütt et al., 2018). Additionally, top-down factors like the observers task (Yarbus et al., 1967) and scene context (Loftus & Mackworth, 1978; Torralba et al., 2006) guide the eyes in scene viewing. Although recent fixation location models perform close to optimal on data from free-viewing experiments (Kümmerer et al., 2016), we have shown in Chapter 2 that even knowing the perfect fixation density map does not produce valid scanpaths, because dynamical aspects play a major role (also see Engbert et al., 2015; Schütt et al., 2017). In Chapter 3 we have shown that the most predictive feature of fixation locations in an image, the central fixation bias, is to a large degree an artifact of the laboratory design. In Chapter 4 and in our paper on the influence of saliency over time (Schütt et al., 2018), we have shown that saliency models, which predict fixations in free-viewing tasks close to what is known as the gold-standard (Kümmerer et al., 2015), predict fixation locations in visual search worse than chance (also see Henderson et al., 2007). Thus, in all chapters limitations of the classical saliency approach have been described, which confirm the need for dynamical models to predict human eye movements in scene viewing. This does not imply that the saliency approach is not important, because it laid the groundwork of further visual attention models. Static saliency modeling is just not a holistic approach for predicting human scanpaths in natural scene perception. To overcome this problem, dynamical models have been formulated in the recent past. All experiments within this thesis were conducted to improve and evaluate the dynamical SceneWalk model of scanpath generation.

The first success for the model was showing that spatial inhibition of return is compatible with our results presented in Chapter 2 and that the SceneWalk model can account for the long-lasting influence of the starting position via the dynamic fixation map. The SceneWalk model generated a characteristic overshoot to the opposite image side which was not present in simulated data of any model without an inhibition of return mechanism.

As a second success for the SceneWalk model, we were able to implement a plausible central fixation bias, which improved the models performance on early fixations of the scanpath. Since the bias explains a large amount of variance in scene-viewing data, it is important to find an appropriate way to model it. To estimate parameters for the model extension we used likelihood based parameter estimation (Schütt et al., 2017).

As a first shortcoming of the model, we found that the systematic distribution of

angles between two adjacent saccades, seen in multiple experiments (Tatler & Vincent, 2008; Smith & Henderson, 2009; Rothkegel et al., 2016), was not yet replicated by the SceneWalk model. By adding a conditional probability map of saccade amplitude and angle, as seen in Chapter 2, the model created an angle distribution close to the empirical distribution (see Fig. 2.12). Thus, the SceneWalk model can be improved rather effortlessly to match experimental data. Unfortunately, the implementation of the distribution of saccadic angles and amplitudes taken from the data does not help to understand why this systematic behavior is manifested in human eye movements. Thus, in a future version of the model a more theoretical approach to model systematic tendencies between adjacent saccades is necessary. An increase in attentional activation of potential saccade targets, which lie in the same direction as the previous saccade, is one possibility to increase the amount of forward saccades. It has been postulated that attention shifts to a new target location before a saccade is executed (Deubel & Schneider, 1996) and that, once the eyes start to move, this attentional shift moves with the eyes (Rolfs, Jonikaitis, Deubel, & Cavanagh, 2011). If this movement of attentional shifts (or so-called remapping) exists, position which lie within the same direction as the previous saccade would indeed be facilitated. Additionally, it has been shown that attention lingers at the previous saccade for a short time-interval, which is compatible with our observation that inhibition of return sets in after a short facilitation of return period (Golomb, Chun, & Mazer, 2008). Both these mechanisms, a short facilitation of return and an increase of attention in the current saccadic direction can be implemented into the SceneWalk model without adding additional maps. The attention map can be asymmetrical to account for the facilitation of forward saccades. The fixation map could start out positive (i.e. producing facilitation of return) and evolve into a negative influence (i.e. inhibition of return) with increasing time.

So far, the SceneWalk model has only been tested on free-viewing and memorization data. In the future, the model will be evaluated on search experiments and other tasks as navigating in real-world environment. For visual search we found an easy way to tune the model to some target characteristics. Human observers adapt their eye movement behavior rather fast to the visibility of the target in the periphery. We believe that this adaptation results from an increase or decrease in the attentional span of searchers, which resembles the size of the attention map of the SceneWalk model. To account for adaptation, the standard deviation of the Gaussian from the attentional map could be adjusted, according to the properties of the target in a visual search task.

The original SceneWalk model uses the ground truth empirical density map as an underlying saliency map, on which a gaze-control model chooses saccade targets. This empirical saliency map is not available when we want to predict scanpaths on any given image. Recently, we published a saliency model (Schütt et al., 2018) based on an image-

computable early-vision model by Schütt and Wichmann (2017). This underlying early vision model accurately modeled data from psychophysical experiments on early vision and thus has proven its psychophysical validity (Schütt & Wichmann, 2017). By combining deep-learning software with receptive field properties of neurons in the early visual cortex, this model performs slightly better than classic saliency models but still performed below the DeepGaze II model (Kümmerer et al., 2016), probably due to the before mentioned aspect that higher-level objects are necessary to predict human viewing behavior in natural scenes. Although the model performs rather well, it's performance could still be improved by incorporating higher-level features and top-down influences. One important outcome of this thesis (Chapter 4) is that we are now able to divide fixations into selective (after a change in saccade direction) and more default fixations (after *saccadic momentum* saccades). For modeling saliency or conspicuity, it is important to know which fixations are placed unintentionally by a default mechanism and which are rather selective. We think that the models performance in predicting fixation locations might be improved by extracting features based only on fixations which are preceded by a large change in saccade direction.

5.3 Final conclusion

This thesis investigated how selected systematic eye-movement tendencies shape human scanpaths in computer based scene-viewing experiments and how these help to increase the predictive power of a dynamical model of saccade generation. In three studies we found systematic eye-movement behavior which goes beyond influences of the image content. The most important implications for dynamical modeling of scanpaths were: (i) Inhibition of return is not a strictly temporal phenomenon in scene viewing but influences spatial selection of saccade targets. The implementation of a dynamic inhibition-of-return mechanism in computational models of saccade generation replicated dynamic progression of human scanpaths. Thus, it seems that inhibitory tagging is a valid driving force for dynamical models. (ii) The initial fixation position and duration have a strong influence on the subsequent scanpath. Although this interacts strongly with the image content, an overshoot to the opposite image side and a long lasting transient were observed independent of the image. (iii) The initial fixation is also primarily accountable for the central fixation bias in scene viewing. Long-lasting initial fixations lead to a significantly weaker central fixation bias than short initial fixations. By experimentally prolonging the initial fixation for 75 ms or more, the central fixations bias was significantly reduced. (iv) Human observers adapt their eye movements very fast when searching for targets of different low-level properties on complex backgrounds. It has been shown before that fixation locations depend on the visual properties of the search target but to our knowledge no work has yet

shown that scanpath properties like saccade amplitudes and fixation durations are influenced by the low-level properties of a search target. (v) Saccades are strongly influenced by the direction of preceding saccades, regardless of the image, subject or task. Although this has been shown before, we were able to show that saccades which maintain direction are less selective than saccades which change direction. We thus interpret saccades which maintain direction as part of a default scanning mechanism. Another indication for the existence of default saccades is that in visual search the visual properties of search targets only influence saccades with a change in direction.

This thesis highlights the important contribution of systematic eye-movement tendencies in scene viewing and scene search and shows that they are an important aspect of eye-movement control. Furthermore, when investigating human eye movement behavior, dynamic aspects of a scanpath need to be considered for a complete understanding of how visual attention is distributed.

References

Abed, F. (1991). Cultural influences on visual scanning patterns. *Journal of Cross-Cultural Psychology*, *22*(4), 525–534.

Anderson, N. C., Donk, M., & Meeter, M. (2016). The influence of a scene preview on eye movement behavior in natural scenes. *Psychonomic Bulletin & Review*, *23*, 1794–1801.

Antes, J. R. (1974). The time course of picture viewing. *Journal of Experimental Psychology*, *103*(1), 62–70.

Arizpe, J., Kravitz, D. J., Yovel, G., & Baker, C. I. (2012). Start position strongly influences fixation patterns during face processing: Difficulties with eye movements as a measure of information use. *PloS one*, *7*(2), e31106.

Baddeley, A., & Turner, R. (n.d.). SPATSTAT: An R package for analyzing spatial point patterns. *Journal of Statistical Software*(6), 1–42.

Bahill, A. T. (1975). Most naturally occurring human saccades have magnitudes of 15 deg or less. *Investigative Ophthalmology*, *14*, 468–469.

Bair, W., & O'keefe, L. P. (1998). The influence of fixational eye movements on the response of neurons in area MT of the macaque. *Visual Neuroscience*, *15*(4), 779–786.

Barthelmé, S., Trukenbrod, H., Engbert, R., & Wichmann, F. (2013). Modeling fixation locations using spatial point processes. *Journal of Vision*, *13*(12), 1:1–34.

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

Bays, P. M., & Husain, M. (2012). Active inhibition and memory promote exploration and search of natural scenes. *Journal of Vision*, *12*(8), 8:1–18.

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

Bickel, P. J., & Doksum, K. A. (1977). *Mathematical statistics: Ideas and concepts*. San Francisco: USA: Holden-Day.

Binaee, K., Diaz, G., Pelz, J., & Phillips, F. (2016). Binocular eye tracking calibration during a virtual ball catching task using head mounted display. In *Proceedings of the ACM Symposium on Applied Perception* (pp. 15–18).

Bindemann, M. (2010). Scene and screen center bias early eye movements in scene viewing. *Vision Research*, *50*(23), 2577–2587.

Bindemann, M., Scheepers, C., Ferguson, H. J., & Burton, A. M. (2010). Face, body, and center of gravity mediate person detection in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(6), 1477–1485.

Blakemore, C., & Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *The Journal of Physiology*, *203*(1), 237–260.

Borji, A., Cheng, M.-M., Jiang, H., & Li, J. (2015). Salient object detection: A benchmark. *IEEE Transactions on Image Processing*, *24*(12), 5706–5722.

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

Borji, A., & Itti, L. (2014). Defending Yarbus: Eye movements reveal observers' task. *Journal of Vision*, *14*(3), 29:1–22.

Borji, A., Sihite, D. N., & Itti, L. (2013). Objects do not predict fixations better than early saliency: A re-analysis of Einhäuser et al.'s data. *Journal of Vision*, *13*(10), 18:1–4.

Botev, Z. I., Grotowski, J. F., & Kroese, D. P. (2010). Kernel density estimation via diffusion. *Annals of Statistics*, *38*(5), 2916-2957.

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

Bruce, N. D., & Tsotsos, J. K. (2009). Saliency, attention, and visual search: An information theoretic approach. *Journal of Vision*, *9*(3), 5:1–24.

Bruce, N. D., Wloka, C., Frosst, N., Rahman, S., & Tsotsos, J. K. (2015). On computational modeling of visual saliency: Examining what's right, and what's left. *Vision Research*, *116*, 95–112.

Buswell, G. T. (1935). *How people look at pictures.* Chicago: University of Chicago Press.

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

Cajar, A., Engbert, R., & Laubrock, J. (2016). Spatial frequency processing in the central and peripheral visual field during scene viewing. *Vision Research*, *127*, 186–197.

Cajar, A., Schneeweiß, P., Engbert, R., & Laubrock, J. (2016). Coupling of attention and saccades when viewing scenes with central and peripheral degradation. *Journal of Vision*, *16*(2), 8:1–19.

Campbell, F. W., & Robson, J. (1968). Application of fourier analysis to the visibility of gratings. *The Journal of Physiology*, *197*(3), 551–566.

Castelhano, M. S., & Henderson, J. M. (2008a). The influence of color on the perception of scene gist. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(3), 660–675.

Castelhano, M. S., & Henderson, J. M. (2008b). Stable individual differences across images in human saccadic eye movements. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *62*(1), 1–14.

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, *9*(3), 6:1–15.

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

Clark, B. (1936). An eye-movement study of stereoscopic vision. *The American Journal of Psychology*, *48*(1), 82–97.

Clark, V. P., Fan, S., & Hillyard, S. A. (1995). Identification of early visual evoked potential generators by retinotopic and topographic analyses. *Human Brain Mapping*, *2*(3), 170–187.

Clarke, A. D., & Tatler, B. W. (2014). Deriving an appropriate baseline for describing fixation behaviour. *Vision Research*, *102*, 41–51.

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The eyelink toolbox: eye tracking with matlab and the psychophysics toolbox. *Behavior Research Methods*, *34*(4), 613–617.

Cornelissen, T. H., & Võ, M. L.-H. (2017). Stuck on semantics: Processing of irrelevant object-scene inconsistencies modulates ongoing gaze behavior. *Attention, Perception, & Psychophysics*, *79*(1), 154–168.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to loftus and masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42–45.

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

De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research*, *52*(4), 317–329.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*(12), 1827–1837.

Deubel, H., Wolf, W., & Hauske, G. (1984). The evaluation of the oculomotor error signal. *Advances in Psychology*, *22*, 55–62.

De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, *22*(5), 545–559.

Dickinson, C. A., & Intraub, H. (2009). Spatial asymmetries in viewing and remembering scenes: Consequences of an attentional bias? *Attention, Perception, & Psychophysics*, *71*(6), 1251–1262.

Di Russo, F., Martínez, A., Sereno, M. I., Pitzalis, S., & Hillyard, S. A. (2002). Cortical sources of the early components of the visual evoked potential. *Human Brain Mapping*, *15*(2), 95–111.

Dodge, R. (1903). Five types of eye movement in the horizontal meridian plane of the field of regard. *American Journal of Physiology–Legacy Content*, *8*(4), 307–329.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433–458.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. New York: Chapman and Hall.

Ehlers, H. (1925). On optically elicited nystagmus. *Acta Ophthalmologica*, *3*(1-2), 254–271.

Einhäuser, W., & Nuthmann, A. (2016). Salient in space, salient in time: Fixation probability predicts fixation duration during natural scene viewing. *Journal of Vision*, *16*(11), 13:1–17.

Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision*, *8*(14), 18:1–26.

Einhäuser, W., Rutishauser, U., Koch, C., et al. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision*, *8*(2), 2:1–19.

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

Engbert, R., & Kliegl, R. (2003a). Microsaccades uncover the orientation of covert attention. *Vision Research*, *43*(9), 1035–1045.

Engbert, R., & Kliegl, R. (2003b). Noise-enhanced performance in reading. *Neurocomputing*, *50*, 473–478.

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., Richter, E. M., & Kliegl, R. (2005). Swift: a dynamical model of saccade generation during reading. *Psychological Review*, *112*(4), 777–813.

Engbert, R., Rothkegel, L. O. M., Backhaus, D., & Trukenbrod, H. A. (2016). Evaluation of velocity-based saccade detection in the SMI-ETG 2W system. Potsdam, Germany: Universität Potsdam, allgemeine und biologische Psychologie.

Engbert, R., Trukenbrod, H. A., Barthelmé, S., & Wichmann, F. A. (2015). Spatial statistics and attentional dynamics in scene viewing. *Journal of Vision*, *15*(1), 14:1–17.

Ferreira, F., Apel, J., & Henderson, J. M. (2008). Taking a new look at looking at nothing. *Trends in Cognitive Sciences*, *12*(11), 405–410.

Findlay, J. M., & Gilchrist, I. D. (1998). Eye guidance and visual search. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 295–312). Oxford, England: Elsevier.

Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing*. Oxford: Oxford University Press.

Findlay, J. M., & Harris, L. R. (1984). Small saccades to double-stepped targets moving in two dimensions. *Advances in Psychology*, *22*, 71–78.

Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences*, *22*(04), 661–674.

Foulsham, T., Gray, A., Nasiopoulos, E., & Kingstone, A. (2013). Leftward biases in picture scanning and line bisection: A gaze-contingent window study. *Vision Research*, *78*, 14–25.

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

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

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, *8*(2), 6:1–17.

Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. *Vision Research*, *51*(7), 771–781.

Geisler, W. S., & Chou, K.-L. (1995). Separation of low-level and high-level factors in complex tasks: visual search. *Psychological Review*, *102*(2), 356–378.

Gilchrist, I. D., & Harvey, M. (2000). Refixation frequency and memory mechanisms in visual search. *Current Biology*, *10*(19), 1209–1212.

Golomb, J. D., Chun, M. M., & Mazer, J. A. (2008). The native coordinate system of spatial attention is retinotopic. *Journal of Neuroscience*, *28*(42), 10654–10662.

Guo, K., Meints, K., Hall, C., Hall, S., & Mills, D. (2009). Left gaze bias in humans, rhesus monkeys and domestic dogs. *Animal Cognition*, *12*(3), 409–418.

Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research*, *18*(10), 1279–1296.

Harel, J., Koch, C., & Perona, P. (2007). Graph-based visual saliency. In *Advances in neural information processing systems* (pp. 545–552). Cambridge, MA: MIT Pres.

Hayes, T. R., & Henderson, J. M. (2017). Scan patterns during real-world scene viewing predict individual differences in cognitive capacityhayes & henderson. *Journal of Vision*, *17*(5), 23:1–17.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*(4), 188–194.

Henderson, J. M. (1992). Visual attention and eye movement control during reading and picture viewing. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 260–283). New York: Springer.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*(11), 498–504.

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

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

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

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, *50*(1), 243–271.

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

Henderson, J. M., & Luke, S. G. (2014). Stable individual differences in saccadic eye movements during reading, pseudoreading, scene viewing, and scene search. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1390–1400.

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

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

Henderson, J. M., Weeks Jr, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental*

Psychology: Human Perception and Performance, *25*(1), 210–228.

Hooge, I. T. C., Over, E. A., van Wezel, R. J., & Frens, M. A. (2005). Inhibition of return is not a foraging facilitator in saccadic search and free viewing. *Vision Research*, *45*(14), 1901–1908.

Ho-Phuoc, T., Guyader, N., Landragin, F., & Guérin-Dugué, A. (2012). When viewing natural scenes, do abnormal colors impact on spatial or temporal parameters of eye movements? *Journal of Vision*, *12*(2), 4:1–13.

Hulleman, J., & Olivers, C. N. (2015). The impending demise of the item in visual search. *Behavioral and Brain Sciences*, *17*, 1–76.

Hwang, A. D., Higgins, E. C., & Pomplun, M. (2009). A model of top-down attentional control during visual search in complex scenes. *Journal of Vision*, *9*(5), 25:1–18.

Ioannidou, F., Hermens, F., & Hodgson, T. (2016). The central bias in day-to-day viewing. *Journal of Eye Movement Research*, *9*(6), 1–13.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10), 1489–1506.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*(3), 194–203.

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

Judd, T., Ehinger, K., Durand, F., & Torralba, A. (2009). Learning to predict where humans look. In *IEEE 12th International Conference on Computer Vision* (pp. 2106–2113).

Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., Hudspeth, A. J., et al. (2000). *Principles of neural science* (Vol. 4). McGraw-hill New York.

Kaspar, K., & König, P. (2011). Overt attention and context factors: the impact of repeated presentations, image type, and individual motivation. *PloS one*, *6*(7), e21719.

Kienzle, W., Franz, M. O., Schölkopf, B., & Wichmann, F. A. (2009). Center-surround patterns emerge as optimal predictors for human saccade targets. *Journal of Vision*, *9*(5), 7:1–15.

Kienzle, W., Wichmann, F. A., Franz, M. O., & Schölkopf, B. (2006). A nonparametric approach to bottom-up visual saliency. In *Advances in neural information processing systems* (pp. 689–696). Cambridge, MA: MIT Pres.

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

Klein, R. (1988). Inhibitory tagging system facilitates visual search. *Nature*, *334*(6181), 430–431.

Klein, R. (2000). Inhibition of return. *Trends in Cognitive Sciences*, *4*(4), 138–147.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., et al. (2007). What's new in psychtoolbox-3. *Perception*, *36*(14), 1–16.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, *4*(4), 219–227.

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

Kootstra, G., de Boer, B., & Schomaker, L. R. (2011). Predicting eye fixations on complex visual stimuli using local symmetry. *Cognitive Computation*, *3*(1), 223–240.

Kowler, E., & Blaser, E. (1995). The accuracy and precision of saccades to small and large targets. *Vision Research*, *35*(12), 1741–1754.

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.

Kümmerer, M., Wallis, T. S., & Bethge, M. (2015). Information-theoretic model comparison unifies saliency metrics. *Proceedings of the National Academy of Sciences*, *112*(52), 16054–16059.

Kümmerer, M., Wallis, T. S., & Bethge, M. (2016). Deepgaze ii: Reading fixations from deep features trained on object recognition. *arXiv preprint arXiv:1610.01563*.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2013). lmertest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). *R package version*, *2*(6).

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, *41*(25), 3559–3565.

Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature*, *369*(6483), 742–744.

Land, M. F., & McLeod, P. (2000). From eye movements to actions: how batsmen hit the ball. *Nature neuroscience*, *3*(12), 1340–1345.

Laubrock, J., Cajar, A., & Engbert, R. (2013). Control of fixation duration during scene viewing by interaction of foveal and peripheral processing. *Journal of Vision*, *13*(12), 11:1–20.

Le Meur, O., & Baccino, T. (2013). Methods for comparing scanpaths and saliency maps: strengths and weaknesses. *Behavior Research Methods*, *45*(1), 251–266.

Le Meur, O., Le Callet, P., Barba, D., & Thoreau, D. (2006). A coherent computational approach to model bottom-up visual attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*(5), 802–817.

Le Meur, O., & Liu, Z. (2015). Saccadic model of eye movements for free-viewing condition. *Vision Research*, *116*, 152-164.

Leventhal, A. (1991). *The neural basis of visual function: Vision and visual dysfunction*

(Vol. 4). Boca Raton, FL: CRC Press.

Lisberger, S. G., Morris, E., & Tychsen, L. (1987). Visual motion processing and sensory-motor integration for smooth pursuit eye movements. *Annual Review of Neuroscience*, *10*(1), 97–129.

Liversedge, S., Gilchrist, I., & Everling, S. (2011). *The oxford handbook of eye movements*. Oxford: Oxford University Press.

Loftus, G. R. (1985). Picture perception: effects of luminance on available information and information-extraction rate. *Journal of Experimental Psychology: General*, *114*(3), 342–356.

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

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*(4), 476–490.

Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley.

Ludwig, C. J., Mildinhall, J. W., & Gilchrist, I. D. (2007). A population coding account for systematic variation in saccadic dead time. *Journal of Neurophysiology*, *97*(1), 795–805.

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

Luke, S. G., Schmidt, J., & Henderson, J. M. (2013). Temporal oculomotor inhibition of return and spatial facilitation of return in a visual encoding task. *Frontiers in Psychology*, *4*, 400.

Luke, S. G., Smith, T. J., Schmidt, J., & Henderson, J. M. (2014). Dissociating temporal inhibition of return and saccadic momentum across multiple eye-movement tasks. *Journal of Vision*, *14*(14), 9:1–12.

Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision*, *9*(11), 8:1–13.

Mannan, S., Ruddock, K., & Wooding, D. (1995). Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-d images. *Spatial Vision*, *9*(3), 363–386.

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision*, *10*(3), 165–188.

Matin, E. (1974). Saccadic suppression: a review and an analysis. *Psychological Bulletin*,

81(12), 899–917.

MATLAB. (2015). *version 8.6.0 (r2015b)*. Natick, Massachusetts: The MathWorks Inc.

Mays, L. E. (1984). Neural control of vergence eye movements: convergence and divergence neurons in midbrain. *Journal of Neurophysiology*, *51*(5), 1091–1108.

Meinecke, C. (1989). Retinal eccentricity and the detection of targets. *Psychological Research*, *51*(3), 107–116.

Mills, M., Hollingworth, A., Van der Stigchel, S., Hoffman, L., & Dodd, M. D. (2011). Examining the influence of task set on eye movements and fixations. *Journal of Vision*, *11*(8), 17:1–15.

Mourant, R. R., & Rockwell, T. H. (1972). Strategies of visual search by novice and experienced drivers. *Human Factors*, *14*(4), 325–335.

Munoz, D. P., & Everling, S. (2004). Look away: the anti-saccade task and the voluntary control of eye movement. *Nature Reviews Neuroscience*, *5*(3), 218-228.

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, *434*(7031), 387–391.

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

Navalpakkam, V., Arbib, M., & Itti, L. (2005). Attention and scene understanding. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of attention* (pp. 197–203). Oxford: Elsevier.

Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, *45*(2), 205–231.

Navalpakkam, V., Koch, C., Rangel, A., & Perona, P. (2010). Optimal reward harvesting in complex perceptual environments. *Proceedings of the National Academy of Sciences*, *107*(11), 5232–5237.

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research*, *46*(5), 614–621.

Noorani, I., & Carpenter, R. (2016). The later model of reaction time and decision. *Neuroscience & Biobehavioral Reviews*, *64*, 229–251.

Noton, D., & Stark, L. (1971). Scanpaths in eye movements during pattern perception. *Science*, *171*(3968), 308–311.

Nuthmann, A. (2014). How do the regions of the visual field contribute to object search in real-world scenes? evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 342–360.

Nuthmann, A. (2017). Fixation durations in scene viewing: Modeling the effects of local image features, oculomotor parameters, and task. *Psychonomic Bulletin & Review*, *24*(2), 370–392.

Nuthmann, A., Einhäuser, W., & Schütz, I. (2017). How well can saliency models predict

fixation selection in scenes beyond central bias? a new approach to model evaluation using generalized linear mixed models. *Frontiers in Human Neuroscience*, *11*, 491.

Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision*, *10*(8), 20:1–19.

Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). Crisp: a computational model of fixation durations in scene viewing. *Psychological Review*, *117*(2), 382–405.

Nyström, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior Research Methods*, *42*(1), 188–204.

Oliva, A. (2005). Gist of the scene. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of attention* (pp. 251–258). Oxford: Elsevier.

Oliva, A., Torralba, A., Castelhano, M. S., & Henderson, J. M. (2003). Top-down control of visual attention in object detection. *IEEE International Conference on Image Processing*, *1*, 253–256.

Ossandón, J. P., Onat, S., & König, P. (2014). Spatial biases in viewing behavior. *Journal of Vsion*, *14*(2), 20:1–26.

Ottes, F. P., Van Gisbergen, J. A., & Eggermont, J. J. (1985). Latency dependence of colour-based target vs nontarget discrimination by the saccadic system. *Vision Research*, *25*(6), 849–862.

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

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, *42*(1), 107–123.

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research*, *45*(18), 2397–2416.

Pointer, J. S., & Hess, R. F. (1989). The contrast sensitivity gradient across the human visual field: With emphasis on the low spatial frequency range. *Vision Research*, *29*(9), 1133–1151.

Polyak, S. L. (1941). *The retina.* Chicago: University of Chicago Press.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance* (Vol. 10, pp. 531–556). Hilldale: Erlbaum.

Posner, M. I., Rafal, R. D., Choate, L. S., & Vaughan, J. (1985). Inhibition of return: Neural basis and function. *Cognitive Neuropsychology*, *2*(3), 211–228.

Privitera, C. M., & Stark, L. W. (2000). Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *IEEE Transactions on Pattern Analysis*

and Machine Intelligence, *22*(9), 970–982.

Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W. C., LaMantia, A.-S., & White, L. E. (1997). *Neuroscience.* Sunderland: Sinauer Associates.

R Core Team. (2014). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org/`

Rayner, K. (1995). Eye movements and cognitive processes in reading, visual search, and scene perception. In J. M. Findlay, R. Walker, & R. W. Kentridge (Eds.), *Eye movement research: Mechanisms, processes and applications* (pp. 3–22). Amsterdam: Elsevier.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 371–422.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506.

Rayner, K., Smith, T. J., Malcolm, G. L., & Henderson, J. M. (2009). Eye movements and visual encoding during scene perception. *Psychological Science*, *20*(1), 6–10.

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.

Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network: Computation in Neural Systems*, *10*(4), 341–350.

Richards, W., & Kaufman, L. (1969). "Center-of-gravity" tendencies for fixations and flow patterns. *Attention, Perception, & Psychophysics*, *5*(2), 81–84.

Riggs, L. A., Ratliff, F., Cornsweet, J. C., & Cornsweet, T. N. (1953). The disappearance of steadily fixated visual test objects. *Journal of the Optical Society of America*, *43*(6), 495–501.

Robson, J., & Graham, N. (1981). Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Research*, *21*(3), 409–418.

Rodrigues, S. T., Vickers, J. N., & Williams, A. M. (2002). Head, eye and arm coordination in table tennis. *Journal of Sports Sciences*, *20*(3), 187–200.

Rolfs, M., Jonikaitis, D., Deubel, H., & Cavanagh, P. (2011). Predictive remapping of attention across eye movements. *Nature Neuroscience*, *14*(2), 252–256.

Rosa, P. J., Gamito, P., Oliveira, J., Morais, D., Pavlovic, M., & Smyth, O. (2015). Show me your eyes! the combined use of eye tracking and virtual reality applications for cognitive assessment. In *Proceedings of the 3rd 2015 workshop on icts for improving patients rehabilitation research techniques* (pp. 135–138).

Rothkegel, L. O. M., Schütt, H. H., Trukenbrod, H. A., Wichmann, F. A., & Engbert, R. (2018). Searchers adjust their eye movement dynamics to the target characteristics in natural scenes. *arXiv preprint arXiv:1802.04069*.

Rothkegel, L. O. M., Trukenbrod, H. A., Schütt, H. H., Wichmann, F. A., & Engbert, R.

(2016). Influence of initial fixation position in scene viewing. *Vision Research*, *129*, 33–49.

Rothkegel, L. O. M., Trukenbrod, H. A., Schütt, H. H., Wichmann, F. A., & Engbert, R. (2017). Temporal evolution of the central fixation bias in scene viewing. *Journal of Vision*, *17*(13), 3:1–18.

Schneider, W. X., & Deubel, H. (1995). Visual attention and saccadic eye movements: Evidence for obligatory and selective spatial coupling. In J. M. Findlay, R. Walker, & R. W. Kentridge (Eds.), *Eye movement research: Mechanisms, processes and applications* (pp. 317–324). Amsterdam: Elsevier.

Schütt, H. H., Rothkegel, L. O. M., Trukenbrod, H. A., Engbert, R., & Wichmann, F. A. (2018). Disentangling top-down vs. bottom-up and low-level vs. high-level influences on eye movements over time. *arXiv preprint arXiv:1803.07352*.

Schütt, H. H., Rothkegel, L. O. M., Trukenbrod, H. A., Reich, S., Wichmann, F. A., & Engbert, R. (2017). Likelihood-based parameter estimation and comparison of dynamical cognitive models. *Psychological Review*, *124*(4), 505-524.

Schütt, H. H., & Wichmann, F. A. (2017). An image-computable psychophysical spatial vision model. *Journal of Vision*, *17*(12), 12:1–35.

Scott, D. W. (2015). *Multivariate density estimation: theory, practice, and visualization*. New York: Wiley.

Shioiri, S. (1993). Postsaccadic processing of the retinal image during picture scanning. *Attention, Perception, & Psychophysics*, *53*(3), 305–314.

Smith, T. J., & Henderson, J. M. (2009). Facilitation of return during scene viewing. *Visual Cognition*, *17*(6-7), 1083–1108.

Smith, T. J., & Henderson, J. M. (2011). Does oculomotor inhibition of return influence fixation probability during scene search? *Attention, Perception, & Psychophysics*, *73*(8), 2384–2398.

Stensola, H., Stensola, T., Solstad, T., Frøland, K., Moser, M.-B., & Moser, E. I. (2012). The entorhinal grid map is discretized. *Nature*, *492*(7427), 72–78.

Stoll, J., Thrun, M., Nuthmann, A., & Einhäuser, W. (2015). Overt attention in natural scenes: objects dominate features. *Vision Research*, *107*, 36–48.

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

Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: effects of scale and time. *Vision Research*, *45*(5), 643–659.

Tatler, B. W., Baddeley, R. J., & Vincent, B. T. (2006). The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task. *Vision Research*, *46*(12), 1857–1862.

Tatler, B. W., Brockmole, J. R., & Carpenter, R. (2017). Latest: A model of saccadic decisions in space and time. *Psychological Review*, *124*(3), 267–300.

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, *11*(5), 5:1–23.

Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, *2*(2), 1–18.

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

't Hart, B. M., Vockeroth, J., Schumann, F., Bartl, K., Schneider, E., König, P., & Einhäuser, W. (2009). Gaze allocation in natural stimuli: comparing free exploration to head-fixed viewing conditions. *Visual Cognition*, *17*, 1132–1158.

Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: Capture of the eyes by new objects. *Psychological Science*, *9*(5), 379–385.

Thiele, A., Henning, P., Kubischik, M., & Hoffmann, K.-P. (2002). Neural mechanisms of saccadic suppression. *Science*, *295*(5564), 2460–2462.

Torralba, A. (2003). Modeling global scene factors in attention. *Journal of the Optical Society of America*, *20*(7), 1407–1418.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological Review*, *113*(4), 766–786.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136.

Trukenbrod, H. A., Barthelmé, S., Wichmann, F. A., & Engbert, R. (2017). Rigorous spatial statistics for gaze patterns in scene viewing: Effects of repeated viewing. *arXiv preprint arXiv:1704.01761*.

Trukenbrod, H. A., & Engbert, R. (2014). Icat: A computational model for the adaptive control of fixation durations. *Psychonomic Bulletin & Review*, *21*(4), 907–934.

Tse, P., Sheinberg, D., & Logothetis, N. (2002). Fixational eye movements are not affected by abrupt onsets that capture attention. *Vision Research*, *42*(13), 1663–1669.

Tseng, P.-H., Carmi, R., Cameron, I. G., Munoz, D. P., & Itti, L. (2009). Quantifying center bias of observers in free viewing of dynamic natural scenes. *Journal of Vision*, *9*(7), 4:1–16.

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (Fourth ed.). New York: Springer.

Vincent, B. T., Baddeley, R., Correani, A., Troscianko, T., & Leonards, U. (2009). Do we look at lights? using mixture modelling to distinguish between low-and high-level factors in natural image viewing. *Visual Cognition*, *17*(6-7), 856–879.

Vitu, F., Kapoula, Z., Lancelin, D., & Lavigne, F. (2004). Eye movements in reading isolated words: Evidence for strong biases towards the center of the screen. *Vision Research*, *44*(3), 321–338.

Võ, M. L.-H., & Henderson, J. M. (2009). Does gravity matter? effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision*, *9*(3), 24:1–15.

Võ, M. L.-H., & Henderson, J. M. (2010). The time course of initial scene processing for eye movement guidance in natural scene search. *Journal of Vision*, *10*(3), 14:1–13.

Wichmann, F. A., Drewes, J., Rosas, P., & Gegenfurtner, K. R. (2010). Animal detection in natural scenes: critical features revisited. *Journal of Vision*, *10*(4), 6:1–27.

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. New York: Springer. Retrieved from `http://ggplot2.org`

Williams, A., & Davids, K. (1998). Visual search strategy, selective attention, and expertise in soccer. *Research Quarterly for Exercise and Sport*, *69*(2), 111–128.

Wilming, N., Betz, T., Kietzmann, T. C., & König, P. (2011). Measures and limits of models of fixation selection. *PLoS One*, *6*(9), e24038.

Wilming, N., Harst, S., Schmidt, N., & König, P. (2013). Saccadic momentum and facilitation of return saccades contribute to an optimal foraging strategy. *PLoS Computational Biology*, *9*(1), e1002871.

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, *5*(6), 495–501.

Yarbus, A. L., Haigh, B., & Riggs, L. A. (1967). *Eye Movements and Vision* (Vol. 2). New York: Plenum Press.

Zelinsky, G. J., Rao, R. P. N., Hayhoe, M. M., & Ballard, D. H. (1997). Eye movements reveal the spatiotemporal dynamics of visual search. *Psychological Science*, *8*(6), 448–453.