Eye Movement during Repeated Viewing of Images

Augenbewegungen bei wiederholter Betrachtung von Bildern

by

Lisa Schwetlick

A thesis submitted to the
University of Potsdam, Germany
in partial fulfillment of the requirements for the degree of

Master of Science
Cognitive Science – Embodied Cognition

Supervisors

Prof. Dr. Ralf Engbert
Dr. Hans Trukenbrod

Department of Biopsychology
University of Potsdam, Germany

June 2018

Abstract

Visual long-term memory in humans has a high capacity and persists over large time spans. In scene viewing studies it has been found to influence eye movements. However, this effect has only been found on relatively short time scales, with images repeated within the same experimental session. The present study investigates the transfer of these effects to longer time scales. The experiment comprised three sessions separated by several days. In the first session participants viewed unfamiliar images. In sessions 2 and 3 the images that were presented could be (a) unfamiliar, (b) familiar, or (c) semantically and structurally similar to previously seen images. Although subjects showed the expected proficiency in recognizing images, we found no or only weak effects of image familiarity on eye movement measures (e.g. fixation durations, saccade amplitudes, central fixation bias). Our results show that the current visual input primarily drives eye movement and long-term image memory from previous sessions only has a weak effect.

Zusammenfassung

Blickbewegungen sind für die Wahrnehmung der visuellen Welt unerlässlich. Nur das Zentrum der Retina, die Fovea, hat eine ausreichend hohe Auflösung, um Details sehen zu können. Die Augen bewegen sich also in einer Abfolge von ruhigen Momenten (Fixationen) und schnellen Sprüngen (Sakkaden) über eine Szene, um verschiedene Aspekte wahrnehmen zu können.

Welche Orte als Fixationsorte gewählt werden ist eine Kernfrage der Blickbewegungsforschung. Sowohl systematische Faktoren, wie die Struktur des okkulomotorischen Systems, als auch inherente Charakteristika des Bildes (Bottom-up) und Zustand und Intentionen des Betrachters (Top-down) spielen dabei eine Rolle.

Das visuelle Langzeitgedächtnis von Menschen kann eine Vielzahl an Bildern über lange Zeiträume hinweg erinnern. Vorhergehende Studien haben gezeigt, dass die Erinerung an ein Bild auch einen messbaren Einfluss auf Blickbewegungen hat. Bei bekannten Bildern sind Sakkaden kürzer und Fixationsdauern länger als bei unbekannten Bildern.

Dieses Phänomen wurde allerdings bisher ausschließlich auf kurzen Zeitskalen untersucht, indem Bilder innerhalb einer Sitzung wiederholt dargeboten wurden. Im Rahmen dieser Arbeit war zu testen, ob sich von dem Effekt von Bildbekanntheit innerhalb einer Sitzung auch auf längere Zeitspannen schließen lässt.

Versuchspersonen haben an drei Sitzungen teilgenommen, die jeweils mindestens einen Tag voneinander entfernt waren. In der ersten Sitzung waren alle Bilder unbekannt. In der zweiten und dritten Sitzung sahen die Versuchspersonen (a) bekannte Bilder, (b) neue, unbekannte Bilder und (c) Bilder die semantische Ähnlichkeit zu bereits gesehenen Bildern hatten. In der dritten Sitzung wurde zusätzlich eine Maske über dem Bild päsentiert, mit der die Versuchspersonen nur im Zentrum des Blickfelds das Bild sehen konnten (die Peripherie war maskiert).

Obwohl die Versuchspersonen die Bilder wie erwartet sehr gut wiedererkennen konnten, konnten wir keine bis nur sehr kleine Effekte der Bildbekanntheit in den Blickbewegungen feststellen. Sakkadenamplituden und Fixationsdauern blieben über die wiederholten Betrachtungen gleich. Auch der Zeitverlauf dieser Maße war über die Bildbedingungen erstaunlich stabil.

Eine weiteres Maß in dieser Studie war wie ähnlich der Blickpfad zu der empirischen Dichte von Fixationspositionen anderer Versuchspersonen auf dem Bild war. Bei diesem Maß der Ähnlichkkeit und bei der mittleren Distanz zum Zentrum des Bildes waren Unterschiede zwischen den Bildbedingungen festzustellen. Das könnte darauf hinweisen, dass Fixationsorte von der Bekanntheit des Bildes auch nach mehreren Tagen noch beeinflusst werden, die Bewegungen an sich allerdings nicht.

Wider Erwarten konnten wir den Effekt von Bildbekanntheit auf Blickbewegungen auf längeren Zeitskalen nicht reproduzieren. Allgemein scheint die Wahl der Fixationsorte und die Ausführung der Blickbewegungen mehr von gegenwärtigem visuellen Input gesteuert zu sein als vom visuellen Langzeitgedächnis.

Acknowledgements

First and foremost, I thank my supervisors Dr. Hans Trukenbrod and Prof. Dr. Ralf Engbert for their valuable guidance and encouragement during the writing of this thesis. Thank you for all the opportunities you have given me and for always taking time to answer my questions, share expertise, and give feedback. This master's thesis and the described experiment were produced in the context of project B05 of the SFB1295 Data Assimilation, funded by the DFG. The creation of the image data set was supported by the Deutsches Technik Museum Berlin, where I was kindly allowed to take to take photographs of the exhibits. I am also very grateful to the staff of the University of Potsdam Eyelab for meticulously conducting the experiment and collecting the data. Lastly, I would like to thank the invaluable people who read drafts of this thesis, supported me, and offered constructive criticism.

Contents

1 Introduction

Eye movement is vital to perceiving the visual world. In order to get an impression of a scene, the eyes move in a sequence of very fast movements called *saccades* and take in information during the stationary moments (*fixations*) between saccades. Saccades and their target locations are influenced by several factors, including tendencies caused by the nature of the environment and oculomotor system, properties of the scene that is being viewed, and the viewer's prior knowledge and intentions. Decoding how exactly fixation locations are chosen and saccades are planned allows insight into the way perception works in the brain.

The main focus of this thesis is the effect of visual long-term memory on eye movement. Humans have a remarkable capacity for remembering a large amount of images over extended periods of time. Previous studies have shown that there are systematic differences in viewing behavior, when participants are familiar with the image they are seeing compared to when the image is unfamiliar. Fixation duration increases and saccade amplitudes decrease with repeated exposure to the image, which indicates that the attentional focus becomes more local. However, currently this effect has only been investigated on a relatively short timescale, i.e. images were repeated within the same experimental session.

In the present study, we investigated the effect image familiarity has on eye movement when images are repeated several days apart. As humans have a very good memory for images, we expected to replicate the previously found image familiarity effects. In order to explore whether visual long-term memory relies on the semantic gist of an image or specific details for recognition, we also included images that were semantically similar to seen images in the stimulus data set.

Unexpectedly, we were not able to reproduce the familiarity effect found in previous studies in the experiment described in this thesis. While participants were able to remember the images, even after multiple days, we found no effects of image familiarity on saccade amplitude and fixation duration when viewing the images. We did, however, find a significant effect of how similar individual gaze paths are to the density of fixation locations generated by other participants; familiar images seemed to evoke a higher conformity in fixation locations with the empirical density. Tentatively, we conclude that the familiarity effect as previously found may not affect eye movement on longer time spans, while qualifying that familiarity does seem to effect the chosen locations for fixation.

Section 2 gives a more detailed overview of the theoretical background and related work on the influences on eye movement. The experimental setup, method and analyses are reported in Section 3. Results are presented in Section 4 and discussed in Section 5. In Section 6 we report our conclusions and suggest a follow-up study to clarify our findings.

2 Theoretical Background

2.1 Eye Movement and Attention

2.1.1 Eye Movement

Humans typically experience the visual world as continuous, in focus and complete. This impression is not immediately apparent from the visual input that the brain receives. When light enters the eye, it creates an image on the light-sensitive receptors (*rods* and *cones*) that constitute the retina. These translate light into nerve impulses which are transmitted to the brain. However, only a small area in the center of the retina, called the *fovea*, relays high resolution information. Photoreceptors which enable color vision, known as cones, are almost exclusively found in the fovea. The periphery of the visual field is several times less sensitive due to a much lower receptor density and the comparatively smaller dedicated area in the visual cortex (Bear, Connors, & Paradiso, 2007; Strasburger, Rentschler, & Juttner, 2011). Peripheral vision, while ineffective at recognizing detail, is useful for basic image information, under low-light conditions and for detecting movement.

In order to get a detailed impression of the surroundings, the fovea must be directed at various parts of the visual world consecutively. To do this, the human eye moves in a succession of short ballistic movements called *saccades*, with intermittent periods of relative stability called *fixations*. A sequence of fixations and saccades is called a *scan path*. Figure 1 shows an example of such a scan path.

Information is registered only during fixations, which are typically between 200 and 400 ms in length depending on the task (Rayner, 1998). During saccades the eye is essentially blind, a phenomenon called *saccadic suppression*, which refers to the finding that vision is impaired during and around the time of a saccade (Matin, 1974). The suppression of visual information is not complete, however, and some things can still be perceived during a saccade. For example, some studies have found that moving sine gratings that appear to be monotonous surfaces when the eye is steady, can be perceived during saccades (Castet & Masson, 2000; Mathôt, Melmi, &

Figure 1: An example of the scan path of one person on one image. The red line segments are the parts of the movement that was identified as a *saccade*. Yellow segments are fixations.

Castet, 2015). Thus, visual perception during a saccade reappears when the image is stabilized on the retina, suggesting that the suppression is caused by blurring due to the fast movement of the saccade.

Similar to the fixation duration, the distance a saccade covers varies with the task. In scene perception saccades cover 4° of visual angle on average, while during reading and typing they only cover 1-2° (Rayner, 1998). The logarithm of the maximum movement speed is linearly correlated with the logarithm of the distance traveled, a principle known as the *Main Sequence* (Collewijn, Erkelens, & Steinman, 1988).

There are a number of other types of eye movement beside saccades and fixations. Pursuit and vergence movements are produced when keeping an object in focus as it moves in the field of view, e.g. a train going by or a ball flying toward a spectator. During fixations the eyes are also not perfectly still. Low amplitude *fixational eye movements* are involved in preventing image fading due to neural adaptation (Martinez-Conde, Macknik, & Hubel, 2004; Martinez-Conde, Macknik, Troncoso, & Dyar, 2006), i.e. a reduction in the firing rate of neurons due to continued exposure to a stimulus. Fixational eye movements comprise tremor movements (*Nystagmus*), slower drift movements, and microsaccades which are fast movements similar to regular saccades but on a smaller scale. There has been some debate as to what extent drift and tremor movements have a useful purpose or whether they are the result of motoric noise and instability in the oculomotor system (Rayner, 1998).

Microsaccades, on the other hand, are likely systematic in nature. A portion of microsaccades directly correct for displacements caused by drifts (Martinez-Conde et al., 2004). Others are not corrective, although their role is not clear. Various studies have found microsaccades to have relations to attentional mechanisms (Engbert & Kliegl, 2003), gaze shifts between closely spaced details (analogous to real saccades), and counteracting retinal fading (Kowler, 2011).

2.1.2 Passive and Active Vision

For a long time the prevailing theory of how we form an impression of the visual world was *passive vision*. According to this account, the brain composes a model of the scene using the snapshots gleaned during fixations while compensating for assumed inadequacies in stimulus perception. This model then serves as a basis for predictions and action planning (Marr, 1982; Aloimonos & Rosenfeld, 1991).

In 1994 Churchland et al. published an alternative view, *active vision*, which proved to be a turning point in the study of visual perception. The authors suggested that there is no such thing as a stable model of the world in the brain. The perception of perfect, detailed, and constant visual input of the entire scene is an illusion brought about by the fact that detailed information is constantly potentially available by simply moving the eyes to focus on the element of interest (Findlay & Gilchrist, 2003; O'Regan, 1992). The idea of an internal model of the world becomes superfluous because the world serves as its own model. Active vision is more a dynamic approach to eye movements and takes into account the physiological realities of the visual system, while it also assigns the periphery of the visual field a critical role in orienting and guiding eye movement. Furthermore, it gives a significant importance to attentional processes.

2.1.3 Attention

The restricted amount of high acuity input to the visual system due to photoreceptors on the retina concentrically decreasing in density around the fovea, is just one manifestation of a more general limitation. The quantity of sensory stimuli at any given moment far exceeds what can consciously be perceived. The inability to perceive all elements of a visual display at once – otherwise "Where's Waldo" would not be a challenge – can equally be observed in all other senses. It follows that there must exist a method of prioritizing certain stimuli over others, of allocating processing resources. The family of mechanisms responsible for this is known as *attention* (Wolfe et al., 2006).

A common example of filtering sensory input is the *cocktail party effect*. At a busy cocktail party, despite there being dozens of auditory streams, most people have no trouble attending to only one (Arons, 1992). Subjects were usually unable to answer questions about the content of unattended auditory streams (Broadbent, 1954; Cherry, 1953), but were sometimes able to react to semantically interesting information in unattended channels (e.g. their own name; Wood & Cowan, 1995).

Visual attention is generally divided into two distinct forms: *overt* and *covert*. Overt attention is observable from the outside: the eyes and head are directed toward the object of interest in order to examine it. Overt attention is most common, as directing the high acuity fovea to the attended object allows the most information to be gained about it. The reverse conclusion, that the locus of attention is likely to be on the fixated element, is a useful assumption in attention research. By recording the movement of the eye using eye tracking technology, it is possible to draw conclusions about visual attention.

Covert attention refers to a shift in attention that is independent of eye and body movement. It is possible to deliberately fixate the eyes on one element and attend to another (for instance when hiding one's intentions from others), but it is far less common and intuitive. Covert attention is harder and less reliable to detect, although some research suggests that microsaccades can imply the direction of covert attention (Engbert & Kliegl, 2003).

2.2 Guidance Principles

Covert attention is not just a mechanism to conceal ulterior motives, however. It plays an important role in the guidance of eye movements. Covert attention shifts to a target usually precede eye movements and thereby the onset of overt attention (Hoffman, 1998). While attention can technically move independently of visual input, it is also a necessary precursor for saccades to new locations. Attention precedes eye movements to a location both when the movement is triggered by internal mechanisms (e.g. attention) and external factors (e.g. something unexpectedly appearing in the periphery; Yantis & Jonides, 1996). In other words, at some point during a fixation, covert attention will shift to another location. The eyes will then follow, reestablishing overt attention.

But what determines which locations attract covert attention, and consequently fixations? Fixation locations in scene viewing are not uniformly distributed over the image. Some areas in an image attract more fixations than others. As the focus of this thesis is on scene viewing, this section will explore the principles that guide attention and eye movements in that field. Eye movements during reading and other activities are guided by somewhat different mechanisms. Guidance principles in scene viewing depend on three factors: (a) the conditions imposed by the environment and the nature of the oculomotor system, (b) the inherent properties of the stimulus image, and (c) the viewer's prior knowledge and intentions.

2.2.1 Systematic Mechanisms

Systematic tendencies are the most general category of mechanisms that guide eye movements. They are independent of the particular image and viewer and stem from the conditions imposed by the visual system, e.g. position of eyes and muscles, and the visual world, e.g. the laboratory setting

(Foulsham & Kingstone, 2012).

Perhaps the most prominent example is the *central fixation bias* (Tatler, 2007). Even some of the first eye tracking experiments (Buswell, 1935) observed the tendency that the center of an image is more frequently fixated than the periphery. The central fixation bias remains remarkably stable over various manipulations. Choosing photographs that omit the photographer's bias (the tendency to arrange photographs so that subject and focus are in the center) does not eliminate the central fixation bias (Tatler, 2007). It also perseveres through starting viewing from a non-central position, and positioning the screen so that its center does not align with the straight ahead position of the eyes (Vitu, Kapoula, Lancelin, & Lavigne, 2004).

In a study by Rothkegel, Trukenbrod, Schütt, Wichmann, and Engbert (2017), the central fixation bias was manipulated by presenting a fixation cross which participants were instructed to fixate until it disappeared. The image then appeared underneath the cross for a given time, while participants "previewed" the image from the location of the cross, before the cross disappeared and they were allowed to freely explore the scene. Forcing participants to preview the image from a predetermined location for 125 ms or more does significantly decrease the central fixation bias (see Figure 2).

The preview benefit indicates that the reason for the bias could be that the center is a good position from which to extract the initial gist of the scene. When forced to extract the gist from another location, the need to begin image viewing in the center is omitted (Rothkegel et al., 2017).

Another systematic bias is the bias for making saccades in the cardinal directions (Foulsham & Kingstone, 2010; Foulsham, Kingstone, & Underwood, 2008). Horizontal saccades are by far the most frequent, followed by vertical saccades. Diagonal saccades are comparatively rare. This effect persists even when the images are square or circular. However, if the image is rotated, so that the horizon appears at an angle, subjects quickly start to make saccades in the cardinal directions relative to the horizon of the image (Foulsham et al., 2008). When subjects are shown images without a clear horizon (such as fractals), they still make predominantly egocentrically horizontal eye movements (Foulsham & Kingstone, 2010), indicating that the bias for cardinal

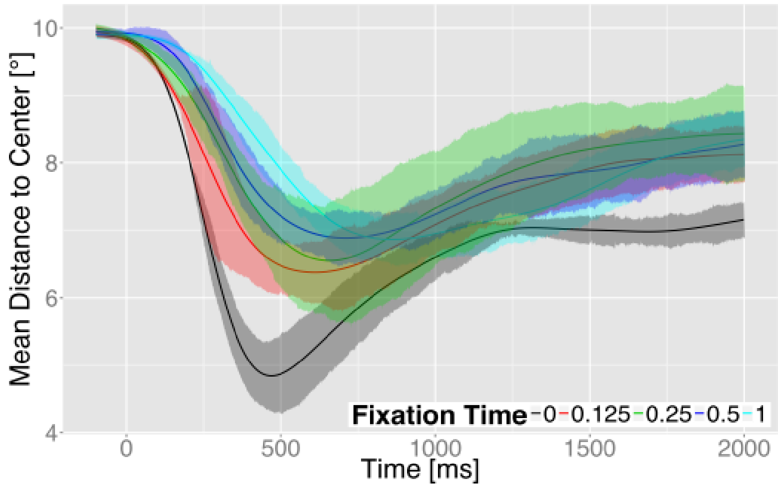

Figure 2: The central fixation bias can be significantly reduced by introducing a preview of as little as 0.125 ms. The distance to the image center 500 ms after the image onset is much smaller, i.e. the bias is larger, when there is no preview than in conditions with a preview (Rothkegel et al., 2017).

directions is an image independent tendency that can be adapted in light of more information.

Lastly, each saccade is not independent from previous saccades. *Inhibition of return*, *facilitation of return*, and *saccadic momentum* are three mechanisms that have been identified to influence the generation of scan paths (Smith & Henderson, 2009).

Inhibition of return was first shown in the field of attention research. A spatial cue is initially beneficial to detecting a target in a location. After a delay however, the speed and accuracy with which a target in the same location can be detected is reduced (Posner & Cohen, 1984). In the field of visual search in scenes, inhibition of return can be interpreted as a mechanism that promotes exploration of a scene (Klein & MacInnes, 1999). Participants tend to not refixate a certain area within some time of the first fixation.

A second mechanism, with the opposite effect, has also been found in scene viewing (Smith & Henderson, 2009). Facilitation of return allows direct returns to the previous (1 back) and second to last (2 back) fixation location.

In the case of the return to the previous location, inhibition of return may cause a delay in saccade execution, but does not decrease the likelihood of such return saccades (Smith & Henderson, 2009).

Saccadic momentum describes a bias in saccade direction, where a saccade is more likely to continue in the same direction as the previous one (Smith & Henderson, 2009). Like inhibition of return, saccadic momentum may also be a tendency that promotes exploration of the scene.

2.2.2 Bottom-up Mechanisms

Bottom-up mechanisms depend critically on the image characteristics of the stimulus image and are independent of the condition of the viewer.

The image statistics in fixated locations differ significantly from non fixated areas (Parkhurst & Niebur, 2003). Fixation locations are correlated with low-level image features such a peaks in luminance, local contrast, edge density, and regions of high spatial frequency content (Mannan, Wooding, & Ruddock, 1996; Reinagel & Zador, 1999; Tatler, Baddeley, & Gilchrist, 2005). While there is no single metric that can predict fixation placement in an image, the distribution of various aggregated image statistics coincides with the distribution of fixations. This discovery has led to a multitude of *saliency models* that aim to predict which areas in an image are likely to be fixated based on image statistics and content (e.g. Parkhurst, Law, & Niebur, 2002; Kümmerer, Wallis, Gatys, & Bethge, 2017). Such saliency models typically produce activation maps that can then be compared to an empirical fixation density map to evaluate its validity (Itti, Koch, & Niebur, 1998). These low-level, bottom-up driven models are supported by the finding that processing in the visual cortex involves mechanisms analogous to extracting image statistics (e.g. neuron clusters responsible for the detection of gratings and contours; Peters, Iyer, Itti, & Koch, 2005).

However, clusters of low-level image features are also indicative of objects, allowing for an alternative explanation for the correlation between low-level features and fixations. As Einhäuser, Spain, and Perona (2008) suggest, attention may instead be driven by object detection, with the coexistence of objects and low-level image statistics as a mere byproduct. The finding that fixations tend to be placed in the center of objects (Nuthmann & Henderson, 2010) supports the idea of object detection as a driving mechanism in the guidance of eye movements.

Incorporating information about object locations into saliency models also makes them more accurate (Kümmerer, Theis, & Bethge, 2014). However, the question of whether eye movements are driven predominantly by object detection or by image statistics remains subject to some debate (Stoll, Thrun, Nuthmann, & Einhäuser, 2015).

2.2.3 Top-down Mechanisms

Previous knowledge, intentions and motivation of the viewer also have an effect on viewing behavior.

Task. When freely performing natural world tasks like making a sandwich, fixation placement ensures continuous execution. Look-ahead fixations are strongly guided by intentions and task requirements and usually directly precede actions (Pelz & Canosa, 2001; Hayhoe & Ballard, 2005; Matthis, Yates, & Hayhoe, 2018).

In the context of scene viewing, early eye tracking research by Yarbus (1967) became seminal evidence of top-down effects on eye movement. Using his own self-invented eye tracking device based on suction cups that attached to the sclera, Yarbus recorded the eye movements of a person given various tasks while viewing images. Figure 3 shows the scan paths elicited by the different viewing instructions. From the recorded gaze paths Yarbus concluded that fixation placement is strongly guided by the viewer's intention.

More recently, concerns have been raised that Yarbus's work is largely anecdotal – counting data from only a single subject – and the methods highly intrusive (DeAngelus & Pelz, 2009; Greene, Liu, & Wolfe, 2012). The subsequent attempts to prove that scan paths can be used to infer the task have only been partially conclusive. DeAngelus and Pelz (2009) did find scan path differences depending on the task, although they were much less striking

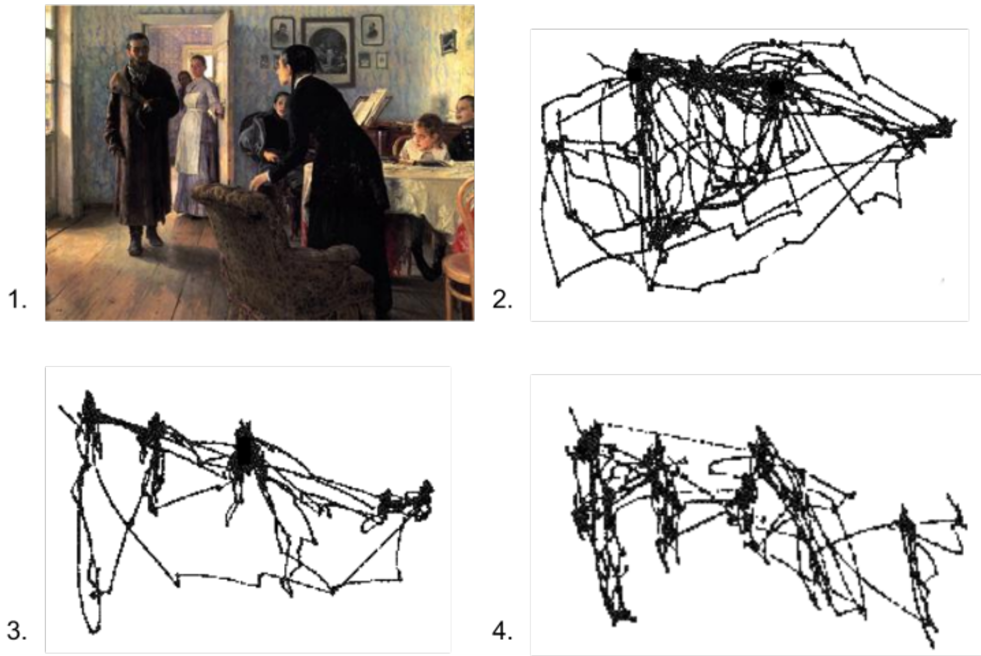

Figure 3: (1) shows the original picture that was shown to the participants in Yarbus's experiment, "The Unexpected Visitor" by Ilya Repin (image taken from `https://commons.wikimedia.org/wiki/File:Ilya_Repin_Unexpected_visitors.jpg`). The other images show the scan paths that the subject's eyes took across the image, depending on the task they had been given (images taken from Yarbus, 1967): (2) free viewing, (3) to guess the ages of the people in the picture and (4) to remember the clothes worn.

than those found by Yarbus. Greene et al. (2012) reported that neither computational pattern classifiers nor human judges were able to accurately identify the task when given a scan path.

Conversely, Borji and Itti (2014), using more advanced machine learning methods, were able to classify tasks at above chance levels, both for the original data generated by Greene et al. as well as their own data set. The extent of task influence on eye movements therefore is an ongoing field of research.

The influence of General World Knowledge. Aside from the effect of task instruction, scene viewing behavior is also influenced by the viewer's knowledge of the world in general and the given scene in particular. For example, as discussed in Section 2.2.1, fixation direction distributions adapt, once knowledge about an image horizon is acquired. Accordingly, saliency models that incorporate contextual priors like a horizon or information about objects (see Section 2.2.2), are more accurate (Torralba, Oliva, Castelhano, & Henderson, 2006; Judd, Ehinger, Durand, & Torralba, 2009).

In visual search tasks, knowledge about the semantics of the target object guides fixations, a phenomenon known as *contextual guidance*. In one experiment Neider and Zelinsky (2006) asked subjects to find target objects in a scene. Objects could appear either in the sky or on the ground and were either semantically scene constrained (e.g. a car is expected to be on the ground; an airship in the sky) or unconstrained (nonsense objects, helicopters). The results showed a benefit of contextual guidance: constrained objects in their expected area were found more effectively, using fewer fixations and in less time. Even the first fixations in the image were guided by context, with more saccades going directly to the expected area. The authors also found that while contextual knowledge biases the search, it does not constrain it; in target absent trials, subjects did check the incongruent regions before giving their final "target absent" answer. For example, they verified that the car was not to be found in the sky before responding.

On a more abstract level, some studies have found that unexpected elements in scenes are fixated earlier, longer and more frequently than expected elements (Henderson, Phillip A., & Hollingworth, 1999; Loftus & Mackworth, 1978). Such studies usually show scenes that include an element that can either be consistent with the context, e.g. a truck in a farm yard, or inconsistent, e.g. an octopus in a farm yard. There has been some debate about this effect (reviewed by Rayner, 2009), especially regarding the limitations of the line drawings used in the earlier studies and controlling for the distinctiveness of the replaced objects. However, recent evidence using realistic pictures or photographs has been in favor of the effect. These studies indicate that regions are fixated earlier when they include unexpected or alarming elements

Neutral Version: Dangerous Version:

Figure 4: Example of an unexpected element manipulation used in the study by Harris et al. (2008). The left image shows a neutral version of the scene (a ball is thrown). The right image shows the dangerous and unexpected version (a baby is thrown).

than in the control scene (example of photorealistic scene that is either neutral or alarming in Figure 4). However, the effect appeared only after several seconds of presentation (Becker, Pashler, & Lubin, 2007; Harris, Kaplan, & Pashler, 2008).

Scene knowledge. Specific knowledge about the scene has also been found to change viewing behavior.

In a study by Castelhano and Henderson (2007) participants had to complete a search task in a gaze contingent window paradigm, i.e. they saw the scene through a small moving window and were unable to gain information from the periphery. The results showed that fewer saccades were needed to find the target when participants were given a valid 250 ms preview of the scene compared with an uninformative or a different scene preview. Participants were only informed of what target they were looking for after the preview. In variations of this experiment, Castelhano and Henderson found that the same benefit exists whether the preview actually includes the target or not and independently of the size of the preview. Previews that showed semantically similar scenes did not produce the benefit. The absence of the benefit in similar previews suggests that the preview provides specific information about features in the scene and not semantic gist information or target location information.

The authors interpret their findings as evidence for an orienting role of the first glimpse of an image. The preview provides information about the scene which remains available over a 2 s delay, and on the basis of which eye movements can be planned.

The interpretation of the first eye movements as an orienting response is consistent with a recent study that found a reduction of the central fixation bias when participants were given a valid 250 ms scene preview compared with a different scene or uninformative preview (Schwetlick, Rothkegel, Trukenbrod, & Engbert, 2017). As discussed in Section 2.2.1, the central fixation bias could be a manifestation of an orienting response: the center of an image is the most informative area, and is therefore the ideal location for the first glimpse. When participants view the image from a predetermined position before exploration, the central fixation bias is reduced (Rothkegel et al., 2017). As subjects are able to gain preview information from the predetermined starting location, the necessity of moving toward the center is decreased.

On a longer timescale, two studies by Kaspar and König investigated eye movements in response to repeated presentations of images, as well as the influence of other top-down and bottom-up processes. Each participant saw the same images 5 times over a single session.

In the first study Kaspar and König (2011a) found that repeated image presentation leads to an increase in fixation duration and a decrease in saccade length and fixation entropy. These results were also influenced by image type (e.g. nature or urban), the subject's motivational state, and their interest in the image material.

The second study by Kaspar and König (2011b) investigated the similarity of fixation locations between viewings. They found that the correlation of fixation locations and image features remained fairly consistent over presentations. Fixation locations in subsequent presentations were moderately correlated with each other, i.e. subjects neither systematically inspected new nor previously inspected regions but a mixture of both. The number of fixated regions decreased over presentations. The authors concluded that the subjects' attentional focus becomes more local with greater image familiarity.

A recent study has corroborated this effect of image familiarity on saccade amplitudes and fixation duration (Trukenbrod, Barthelmé, Wichmann, & Engbert, 2017). The authors also applied spatial statistics to the results and found that points are more strongly spatially correlated within 4° in the second presentation than in the first, using the pair correlation function. In other words, fixations clustered more strongly in the second presentation, with the probability of a fixations appearing closer together larger in the second presentation. The authors conclude that their results are consistent with the idea of an increasingly local attentional focus.

2.2.4 Time Dynamic

Much of eye movement research has centered around the analysis of fixation and saccade characteristics independently of when they occurred in the trial. However, factors that influence eye movement behavior, like scene familiarity, do change over time.

As discussed in Section 2.2.3, the orienting hypothesis suggests that the first glimpse of an image is crucial to later movement planning. This initially acquired representation is capable of guiding subsequent eye movements (Castelhano & Henderson, 2007). It is still unclear, however, how this representation changes over time when the entire image is available at each fixation.

In the area of visual search, studies have found evidence of a coarse to fine dynamic, with long saccades becoming shorter and short fixations becoming longer over the course of a trial (Over, Hooge, Vlaskamp, & Erkelens, 2007). A strategy of moving from general to detailed perception in visual search is a plausible explanation for the observed effect. However, Godwin, Reichle, and Menneer (2014) argue that the effect may not be a strategy but instead arises as an emergent property caused by other guiding principles. Their suggested model comprises four principles: (a) bias in favor of spatially closer targets,

(b) inhibition of return, (c) bias against dissimilar stimuli, and (d) a relation of fixation duration and following saccade length. The authors provide a simulation that reproduces the coarse to fine dynamic found in empirical data.

In scene viewing experiments a similar effect can be found. Antes (1974) found an increase in fixation duration and a decrease in saccade amplitude over time in a trial. Antes (1974) also reported the mean informativeness of the fixated area by using previously given ratings of grid fields in each image and found that the informativeness decreases over time. In a more recent study, Mills, Hollingworth, van der Stigchel, Hoffman, and Dodd (2011) investigated changes in saccade amplitude and fixation duration with regard to given task instructions. The results show that fixation duration increases over time, with the duration increasing more in free viewing task than in a memory or search task. The development of saccade amplitudes over time was similar over tasks, with a rise in amplitude at the beginning and an insignificant decrease later in the trial. Castelhano, Mack, and Henderson (2009) also found a significant increase in fixation duration over fixation number, but only in the first five fixations, after which it largely remained constant. The central fixation bias is another measure that can be found to influence viewing behavior strongly in the beginning of a trial, and less as time goes on (Tatler, 2007; Rothkegel et al., 2017).

Other studies have tried to quantify the influence of top-down and bottom-up processes on eye movement behavior over time. However, little agreement can be found as research methods differ significantly between studies. One study found that influence of bottom-up factors was strongest at the beginning of a trial (Parkhurst et al., 2002). The authors determined the influence of bottom-up factors by computing a bottom-up saliency map and then evaluating it at the fixation positions. Conversely, Tatler et al. (2005), using a different metric, found no change in the influence of image features over time. Tatler et al.'s metric compares the image features in fixated regions with non-fixated regions while correcting for the central fixation bias. The authors argue that Parkhurst et al.'s results were entirely due to the central fixation bias, as Tatler et al. found the same results when not correcting for it. The results also showed that consistency in fixation locations between observers decreased over time.

In terms of modelling of eye movement, much work has been done on developing saliency maps and predictions of fixation densities. Saliency models calculate static activation maps for images, which can successfully predict where people fixate when all datapoints are aggregated over time (Bylinskii et al., n.d.). However, as they are static, they do not account for differences in fixation behavior over time.

More recently, dynamical models which produce scan paths have gained traction, e.g. LeMeur and Liu (2015) or Engbert, Trukenbrod, Barthelme, and Wichmann (2015). The *SceneWalk* model (Engbert, Trukenbrod, et al., 2015) dynamically generates fixation locations, dependent on (a) an underlying density, (b) a local activation dependent on the current fixation location, and (c) an inhibition of return mechanism. These generated scan paths simulate empirical fixations successfully along a number of metrics such as saccade distribution and spatial correlations.

2.3 Visual Memory

Visual long-term memory is the representation of an image "that has not been continuously actively held in mind" (Brady, Konkle, & Alvarez, 2011). Under this definition it includes any time span between seconds and months or years. In order to estimate memory capacity, two factors have to be considered: number of items and fidelity.

Studies have consistently shown that humans can remember a remarkable number of images over significant time periods. Shepard (1967) showed participants 600 images, the memory of which was tested by a two-alternative forced choice test for 4 months, 1 week, 3 days, or immediately after memorizing. Mean recognition scores decreased with delay but were remarkably high even after a week (87%). After 4 months participants still performed above chance level. In another study (Standing, 1973) participants memorized 10000 images in 5 sessions of 2000 images each. In a memory test directly after the fifth session, participants remembered an average of 83% of the presented subset of images. These capacity studies have been criticized for using images from distinctive categories, i.e. there were no images that were semantically similar, so participants only needed to remember the gist and not the details of a scene.

The question of how much detail of a scene is stored in visual memory has attracted a lot of discussion in literature. On a short timescale the change blindness effect indicates that representations are very general and lacking in detail. The change blindness effect is demonstrated by presenting subjects with images of objects or scenes that change in some way, with the transition between states masked. Participants are usually oblivious to small changes, as long as they do not stand out or change the gist of the scene (Simons & Levin, 1997). Extrapolating from the change blindness effect to long-term visual memory would mean that the memory representation of the scene is not detailed but contains mostly the gist.

However, more recent studies have shown that visual memory does contain more detail, and that subjects are able to tell the difference between two instances of a similar object if the object was previously attended to (Hollingworth, 2006; Simons & Rensink, 2005). In one study subjects were shown 2500 semantically distinct objects (Brady, Konkle, Alvarez, & Oliva, 2008). They reliably remembered which instance of an object they had seen and even in which state they had seen it, e.g. whether the cup was empty or full.

Memory for images and image details accumulates over viewing time (Melcher, 2006). Participants' performance in answering questions about a picture, e.g. on color, location and objects present, increased linearly with the total amount of time they had viewed the picture.

Visual long-term memory has a high capacity and has been shown to include detailed information about the scene. Repeated image presentations can therefore be used to study the effects of scene familiarity on eye movement (Kaspar & König, 2011b, 2011a).

2.4 Approach

In this thesis we explore the effects of visual long-term memory on eye movement measures. Studies have shown visual long-term memory is detailed and has a high capacity over large time delays (Section 2.3). Related work in this area of research (e.g. Kaspar & König, 2011b, 2011a; Trukenbrod et al., 2017) has shown that image familiarity has an effect on viewing behavior. However, effect of image familiarity has only been tested on short timescales of visual long-term memory: images were repeated within the same session, i.e. $<$45 minutes apart.

It has not been previously tested whether the familiarity effect on eye movement persists over several days, as the memory for the image does. The presented experiment addresses this gap in the research. A persistence of the effect over long time spans would indicate that it is indeed a direct result of image familiarity. If the effect is uncoupled from image memory, i.e. it disappears while memory persists, it is likely to be caused by different mechanisms.

In this experiment, much like in the studies done by Kaspar and König (2011b, 2011a) and Trukenbrod et al. (2017) participants saw the same images repeatedly. In contrast to these earlier studies, however, the experiment took place over three sessions on separate days, with images repeated several days apart. As an additional comparison we also included images that were semantically similar. Having a factor for semantically similar images allows us to control for a possible influence of gist memory as opposed to specific scene memory and also adds complexity to the recognition task.

While the first two sessions employ a regular free scene viewing paradigm comparable to the methods used in previous studies, in the third session participants inspect the images with a superimposed gaze contingent moving window. Restricting peripheral vision enables us to study the influence of visual long-term memory on fixation guidance when the current perception of the image is severely limited.

In order to define the scope and focus of this thesis, we defined the following research questions (RQ):

RQ1 Repeated Viewing: Does the gaze path elicited by an unfamiliar image differ systematically from gaze paths on familiar images (seen previously), when several days pass between presentations? Specifically, how are fixation duration, saccade amplitude, and central fixation bias affected?

As image memory persists over long time spans, we expect to find the same tendencies as found in previous, shorter timescale studies, namely evidence for an increasingly local attentional focus evidenced by increasing fixation duration and decreasing saccade amplitudes. We also expect a reduced central fixation bias in familiar scenes.

RQ2 Conformity with Fixation Density: Do the regions that are fixated systematically differ between the familiar and unfamiliar Image Conditions?

In order to explore whether scene familiarity has an effect on fixation placement in the scene, we measure how similar individual gaze paths are to the empirical fixation density of the first viewing of the image. We expect to find a difference between the conditions, but do not have concrete ideas whether familiarity will make gaze paths more or less similar to the empirical density.

RQ3 Semantic Similarity: Is the effect of familiarity on gaze paths specific to the image or the semantic gist of the image? This question is exploratory and addresses how visual image information is remembered. If image memory mostly relies on gist information, gaze paths on semantically similar to seen images will resemble those on familiar images. Correspondingly, if image memory relies on image specific traits, gaze paths on semantically similar to seen images will be more comparable with gaze paths on new images.

RQ4 Time Dynamic: How do the measures from RQ1 and RQ2 develop over the viewing time within a trial?

As discussed in Section 2.2.4, there is some evidence of a coarse to fine viewing strategy in scene viewing. We explore how eye movement measures in the first and second presentation differ over time to gain information about which factors influence eye movements at various times in a trial. A closer look at the time course of scene viewing may allow conclusions about how soon memorized image information is available.

RQ5 **Guidance when information is limited**: In the absence of peripheral information, do participants use their knowledge of an image to guide eye movements?

If visual long-term memory works similarly to a image preview, the results of the third session should be similar to the findings of Castelhano and Henderson (2007). In contrast to that study however, in the present study, no visual search task was given. Nonetheless, we expect participants to exhibit more goal oriented behavior to be triggered by memory in the "familiar" condition. Goal-oriented viewing should manifest itself as an increase in saccade amplitude and fixation duration, and an increase in conformity with the empirical fixation density.

3 Method

The experiment took place over three sessions spread over several days. In the first session participants explored unfamiliar images with the instruction to memorize them.

In the second session they were shown a set of images that contains images from the following categories:

1. seen previously in Session 1
2. semantically similar to images in Session 1
3. never seen before.

At the end of Session 1 and 2 participants were given a memory test to ensure active participation.

In Session 3 participants were presented with images obscured by a gaze-contingent moving window mask. They see images from four different categories:

1. seen twice, in Session 1 and Session 2
2. seen once in Session 1
3. seen once in Session 1 and seen a similar image in Session 2
4. never seen before.

After each image in Session 3, participants indicated whether they had ever seen it before.

3.1 Image Data Set

For the present experiment we generated a new set of stimulus images. The set consisted of 188 pairs of two semantically related photographs (376 photographs in total). Each photograph was taken using a DSLR camera. We aimed to fulfill two criteria.

Firstly, the entire image should be in focus, as focus in an image guides eye movements (Tatler, 2007). Secondly, the image should not contain text. Reading induces very distinctive eye movements (Rayner, 2009) and is therefore unsuitable to a scene exploration task.

Not all photographs fulfill these criteria completely. 26 images include text or letters in a way that was not possible to be remedied by cropping. These are mostly cases where an object in the photograph has a brand name. The "in focus" criterion is more fluid and must be judged individually for every image. After manual inspection, some images we judged as viable although they showed areas that are not perfectly in focus if they were small or did not significantly distract from the image contents.

Lastly, generating semantically similar pairs introduced its own set of constraints. Both images in a semantic pair must show a similar scene, taken with similar, if not identical camera settings. The two photographs must show two different instances of a similar scene, e.g. two different kitchen cupboards or two different roads. The same scene from different angles or the same scene with the same objects in different positions was not considered a valid pair. Within the pairs focal length, ISO value and F value were kept constant where possible. The images where these values differed significantly were manually reviewed and deemed acceptable.

The image conversion and cropping was done with the open source program Rawtherapee (Horváth, 2017). Where possible we cropped the images to minimize the photographer's bias.

3.2 Setup

The images were presented at a resolution of 1500x1000 px on a ViewPixx 3D screen. At a distance of 60 cm from the monitor images subtended 48° by 28° of visual angle. The ViewPixx 3D screen is specifically designed for vision research and has a diagonal size of 61.4 cm, a aspect ratio of 16:10 (1920x1080 px), and a refresh rate of 100-120 Hz. Participant answers were logged using a ResponsePixx button box (Figure 6). This handheld device reduces movement when responding compared to a conventional keyboard.

Figure 5: Some examples of the semantic pairs of images used in the experiment.

The experiment code was written using Matlab (MathWorks Inc., 2017) and the PsychToolbox (Brainard, 1997).

3.3 Eye Movement Recording

Eye movements were recorded using an Eyelink 1000 video-based, desktop-mounted eye tracker, recording at a sampling rate of 1000 Hz. It was calibrated using a 9-point target grid and validated with the same method. If

Figure 6: The button box that was used to log participant responses in the recall task (Image taken from `http://vpixx.com/wp-content/uploads/2014/07/RESPONSEPixx_1.jpg`).

validation failed, the calibration was re-initiated. The eye tracker provides monocular data and we tracked each participant's dominant eye. During the experiment we recalibrated the device every 14 trials.

3.4 Participants

Data collected from 32 people were included in the analysis of the experiment. The group consisted of 28 females and 4 males between the ages of 18 to 49. They all demonstrated normal or corrected to normal vision. Participation was rewarded with €8/h or course credits. Participants were able to earn an additional reward of max. €2 in the memory test in the first and second session and €4 in the third session by correctly identifying images.

3.5 Procedure

Each participant in the experiment took part in 3 separate sessions (summarized in Figure 7). These took place over the course of several days and never on the same day. Most participants completed all sessions within 7 days, however some participants took longer (21 days at most).

In each session participants were asked to blink as little as possible during the trials. In Session 1 and 2 each image was preceded by a fixation cross and participants were asked to fixate the cross until it disappeared. The

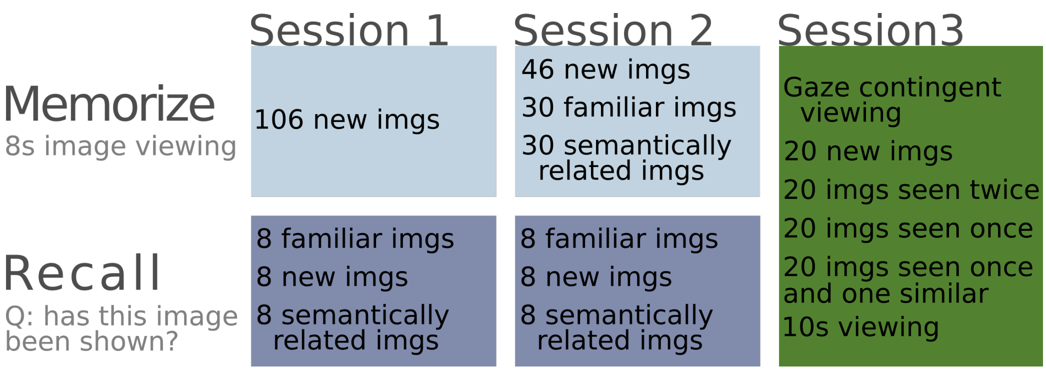

Figure 7: Summary of the experiment design.

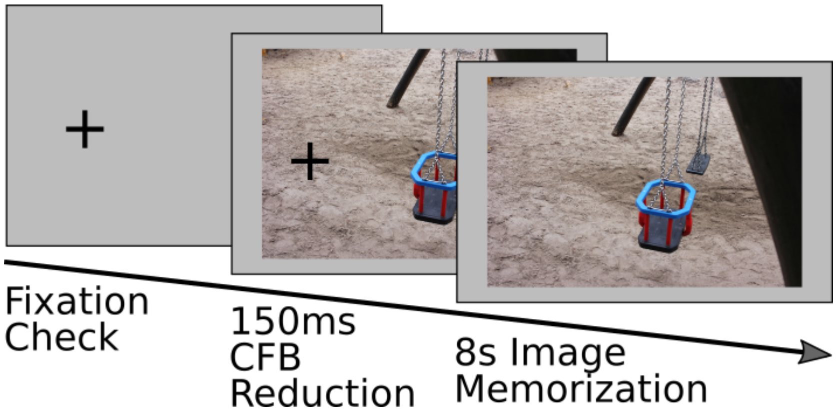

Figure 8: Experiment procedure of presenting the images in Session 1 and 2.

image appeared underneath the cross during fixation for 150 ms before the cross disappeared, see Figure 8. This technique has been shown to reduce the central fixation bias (Rothkegel et al., 2017).

Session 1. Participants were shown 106 new images for 8 s each and are asked to memorize them. The session ended with a memory test of 24 trials, with 8 pictures each in the categories "new", "similar to seen images", and "seen". Participants had to press a button to indicate whether they knew the picture or not.

Session 2. As in Session 1, participants were shown 106 images for 8 s each and are asked to memorize them. These images include 30 images seen in Session 1, 30 images that were semantically similar to images seen in Session 1, and 46 new images (16 were used only for the memory test at the end of the session). The session ended with a memory test of 24 trials: 8 pictures were new, 8 were similar to seen images seen previously in the session, and 8 had been seen previously in the session.

Session 3. Participants were shown 80 images for 10 s each with a superimposed gaze contingent moving window. The mask is a version of the image scrambled in Fourier space to maintain the same image frequencies without revealing any semantics (Figure 9), as suggested by Einhäuser and Nuthmann (2016).

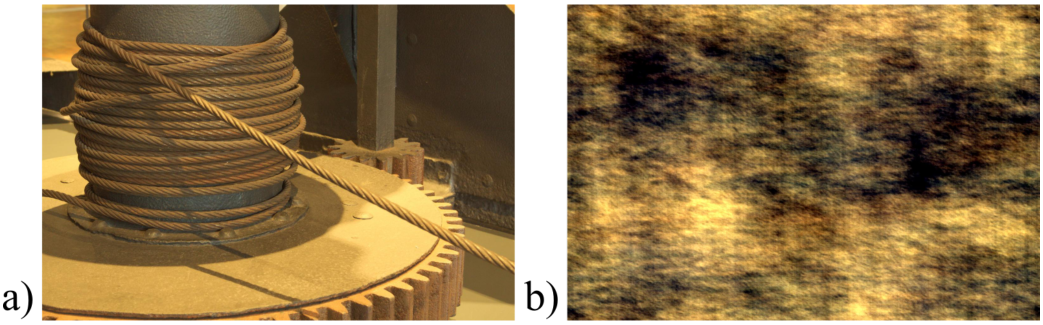

Figure 9: An example of a scrambled image. a) shows the original image; b) shows a version which has been phase scrambled.

The scrambling is done by Fourier-transforming the image and then multiplying the values f_{ij} with a uniformly distributed random value r:

$$f_{jk} \cdot e^{i \cdot r_{jk}}.$$

As the matrix of Fourier values is symmetric, each value pair needs to be assigned complementary phase shifts.

After the gaze contingent viewing, participants were asked to indicate, using the button box, whether the image is familiar. The images include:

1. 20 images seen in Session 1
2. 20 images seen in Session 1, the semantic partner of which was seen in Session 2
3. 20 images seen in both Session 1 and 2
4. 20 new images.

3.6 Analysis

Saccade Detection. For the data analysis we used the programming language R (R Core Team, 2017). Saccades and fixations were detected using the "Microsaccade Toolbox" (Engbert, Sinn, Mergenthaler, & Trukenbrod, 2015). Datapoints that were identified as blinks were excluded from the analyses.

Conformity Metric. In order to measure whether participants fixated the same regions as other participants given that they knew the image or not, we computed the log-likelihood of each fixation location given the empirical fixation density of that image. The described procedure was suggested by Kümmerer, Wallis, and Bethge (2015) and Schütt et al. (2017) for evaluating the performance of fixation models. In the present experiment we use this metric as a measures the conformity of a given gaze position with the empirical density of fixations on the image. It can be compared across the experimental conditions to explore how image familiarity influences fixation placement relative to important or salient areas in the image.

The conformity metric for a given fixation x is computed using the following steps. First, we generate a density map d of all fixations on the same image, using first presentation trials only and excluding the gaze path that is being evaluated. The density map was calculated on a 128x128 grid using the spatstat package for r and a bw.ppl bandwidth (Baddeley, Rubak, & Turner, 2015). d is normalized to be a probability density function, i.e. the area under the density equals 1 and values can be interpreted as relative likelihoods.

We then evaluated the density for each fixation x with coordinates x_1, x_2, by getting the density value at each fixation location and taking the logarithm.

$$conformity(x) = \log 2\big(d(x_1, x_2)\big)$$

This results in the log-likelihood difference of the evaluated gaze path and the empirical density. In the null model, a uniform distribution, the fixation probability of each grid point is 2^{-14}. Subtracting -14 from the each conformity value therefore gives the difference of the likelihood of the fixations given the null model and the likelihood of the fixations given the empirical density. The conformity value is a metric in bit/fix of how much better the empirical density is at predicting the fixation data than a random guess.

Mixed Linear Models. For the metrics saccade amplitude, fixation duration, distance to center, and conformity we used mixed linear models to find systematic differences between the different conditions. To compute the linear mixed models we used R (R Core Team, 2017) and the packages lme4 (Bates, Mchler, Bolker, & Walker, 2015) and RePsychLing (Baayen, Bates, Kliegl, & Vasishth, 2015).

A mixed linear model of experimental data comprises fixed and random effects. The question of how to find a suitable structure, particularly for the random effects has been subject to some debate. While Barr, Levy, Scheepers, and Tily (2013) advocated using maximal random effects structures to increase generalizability, Bates, Kliegl, Vasishth, and Baayen (2015) argue that real data often cannot support complex maximal models. Trying to fit a model with more parameters than it can support leads to models that are uninterpretable due to overparametrization.

We used the method described by Bates, Kliegl, et al. (2015) in order to find a suitable random effects structure for each model. They advocate iteratively reducing the maximal model by removing factors that do not contribute meaningfully to explaining variance in the data. A tool for identifying when components can and should be removed is the random effects Principal Component Analysis. The reduced models can then be compared to the more complex models in order to gauge whether valuable information was lost using an anova.

In order to ensure all assumptions of mixed linear models were met, we inspected the residuals plots. When there was a deviation from the normality or homoscedasticity assumptions, we applied the Box-Cox power transform using the RePsychLing package (Baayen et al., 2015).

In the linear mixed effects models, we used treatment contrasts to compare the fixed effects with each other. A t-value of over 2 was taken to indicate significance.

Confidence Intervals. As the study used a within-subject design, regular standard deviations have to be interpreted under the caveat that there may be a strong between-subject variance. Even when there is not much variation within subjects, if between-subject variation is high, standard deviations can be increased, obscuring potentially systematic effects.

In order to make our results more interpretable, we present them using standard deviations and confidence intervals, as suggested by Cousineau (2005). In Cousineau's method, the subject mean is subtracted and the group mean is added to each data point. Let x_{ij} be a data point from the i^{th} participant in the j^{th} condition ($i = 1, ..., N; j = 1, ..., M$):

$$y_{ij} = x_{ij} - \frac{\sum_{l=1}^{N} x_{il}}{N} + \frac{\sum_{k=1}^{N} \sum_{l=1}^{M} x_{kl}}{NM}$$

This procedure effectively removes interindividual variation. We then applied the correction proposed by Morey (2008), which makes the measure of variance correspond in size to those based on the anova mean squared error. The corrected standard deviation is obtained by multiplying the standard deviation of y_{ij} with the correction factor $\sqrt{\frac{M}{M-1}}$.

In order to obtain 95% confidence intervals from the corrected standard deviation, we multiply it with the value of the t-distribution at 0.95 using the appropriate degrees of freedom.

4 Results

The measurements fixation duration (Section 4.1), saccade amplitude (Section 4.2), distance to image center (Section 4.3) and conformity (Section 4.4) were compared over the sessions (1, 2, 3) and conditions that the images could appear in ("new", "double", "similar" for Session 1 and 2; "seen twice", "seen similar", "seen once", "new" for Session 3). For each group of measurements, we performed a mixed linear effects analysis comparing Session 1 and 2, and a second one for Session 3. As the viewing paradigm in Session 3 differed greatly from Session 1 and 2, direct comparisons would not be appropriate. In the first analysis we are especially interested in the interaction between the image conditions and the session, as this will indicate whether familiar images elicit different fixation durations than unfamiliar images.

4.1 Fixation Duration

At first inspection, there were only small visible differences in fixation duration between the sessions and conditions (see Figure 10).

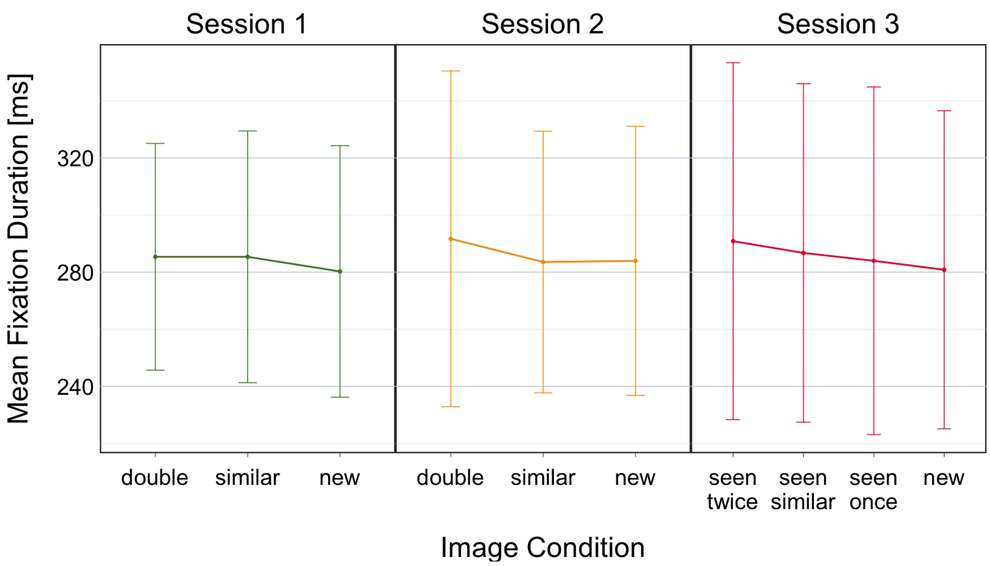

Figure 10: Fixation duration over conditions and sessions. The error bars represent 95% confidence intervals, computed as stated in Section 3.6.

We performed a linear mixed effects analysis as described in Section 3.6 of the fixation duration over the conditions in the first two sessions. Session, Image Condition, and their interaction were entered as fixed effects. We entered Subject and Image as random effect intercepts as well as a by-Subject slope of Session.

```
FixationDuration ~ Session * ImageCondition +
    (1+Session|Subject) + (1|Image)
```

In order to meet the assumptions of homoscedasticity and normality we transformed the fixation duration measurements by raising them to the power of 0.06. Visual inspection of residual plots did not reveal any obvious deviations.

In this analysis, only the main effect of Image Condition "new" versus "double" was significant (see Table 1), i.e. images in the "new" condition were fixated shorter than images in "double" condition.

In a second linear mixed effects analysis we investigated the effects of image familiarity in the third session. We used fixation duration values raised

Fixed effects			
	Estimate	Std. Error	t value
Intercept	1.3968	0.0016	879.87
Session 2 vs. 1	0.0007	0.0008	0.91
Image Condition "similar" vs. "double"	-0.0001	0.0004	-0.33
Image Condition "new" vs. "double"	-0.0008	0.0004	-2.03
Session 2 vs. 1 : Image Condition "similar" vs. "double"	-0.001	0.0007	-1.55
Session 2 vs. 1 : Image Condition "new" vs. "double"	-0.0003	0.0006	-0.56

Table 1: Fixed effects results in the mixed linear model of fixation duration in Session 2 compared with Session 1, and Image Conditions "similar" and "new" compared with "double".

Fixed effects			
	Estimate	Std. Error	t value
Intercept	1.7538	0.0041	423.08
Image Condition “similar” vs. “twice”	0.0007	0.0009	0.82
Image Condition “once” vs. “twice”	-0.0003	0.0008	-0.39
Image Condition “new” vs. “twice”	<0.0001	0.0009	-0.05

Table 2: Fixed effects results in the mixed linear model of fixation duration in Session 3, comparing the Image Conditions “twice”, “similar”, “once” and “new”. All factors are compared the Image Condition “twice”.

to the power of 0.1, and entered Image Condition as a fixed effect, Subject and Image as random intercepts and a by-Image slope of Image Condition.

```
FixationDuration ~ ImageCondition + (1|Subject) +
    (1+ImageCondition|Image)
```

We found no significant effects of the image familiarity on fixation duration in Session 3 (see Table 2).

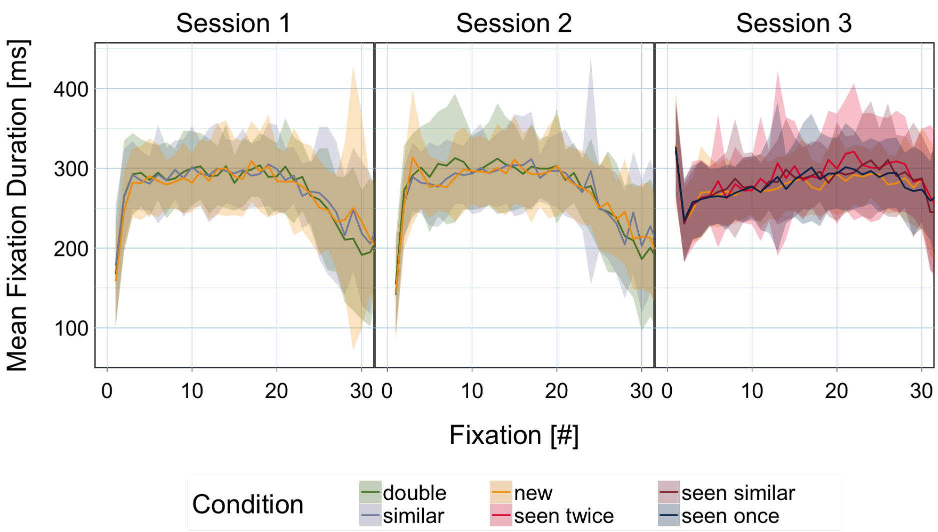

Figure 11: Fixation duration over the fixation number within the trial. The ribbons represent standard deviations, computed using the Cosineau method as stated in Section 3.6.

We inspected changes in fixation duration over the course of a trial to see whether a possible effect may be confined to one part of the trial. In line with the findings from the mixed linear model, plotting the fixation duration by condition and fixation number (Fig. 11) did not reveal any differences of fixation duration over time. The drop-off in mean fixation duration in Session 1 and 2 after saccade 20 is likely caused by data sparsity. Only participants who make short saccades will even reach 20 saccades in one trial.

As we were unable to find any effects of image familiarity on fixation duration, we considered the possibility that the image familiarity effect found in other experiments may be confounded with an effect of the trial number in the session. The second presentation of an image necessarily occurs later in a session than the first. Therefore, an effect of trial number would affect the second presentation more than the first. Figure 12 shows that no such fatigue effect was found. The peak at the beginning of Session 1 can likely be disregarded as the participants were getting used to their environment and the task.

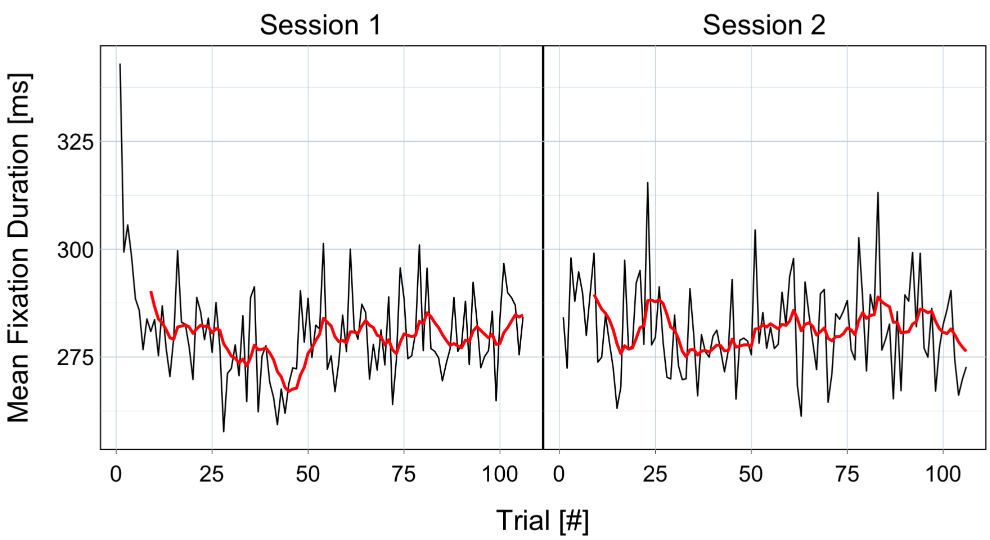

Figure 12: Fixation duration over ordinal trial number. The red line is a smoothing of the data using a causal kernel rolling weighted mean.

4.2 Saccade Amplitude

Figure 13 shows that the saccade amplitude in Session 3 was much lower than in the first two sessions. This difference is expected, as participants typically make shorter saccades during the gaze contingent moving window manipulation. Otherwise differences between conditions were small.

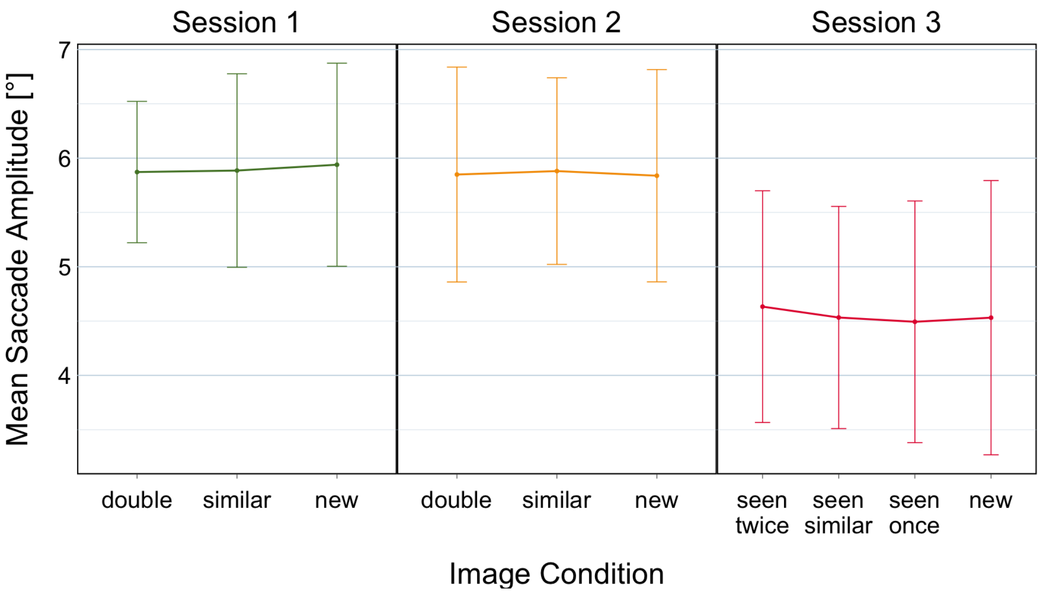

Figure 13: Saccade amplitude over Session and Image Condition. The error bars represent 95% confidence intervals, computed as stated in Section 3.6.

We modeled the data using linear mixed models according to the procedure described in Section 3.6. In this model the fixed effects were Session, Image Condition and their interaction. Subject and Image were included as random effect intercepts as well as a by-Subject slope of Session and a by-Image slope of Image Condition.

```
Amplitude ~ Session * ImageCondition +
    (1+Session|Subject) + (1+ ImageCondition|Image)
```

The assumptions of homoscedasticity and normality were met by raising the measurements to the power of 0.22. Visual inspection of residual plots

Fixed effects			
	Estimate	Std. Error	t value
Intercept	1.4096	0.0075	187.85
Session 2 vs. 1	-0.0031	0.0048	-0.64
Image Condition “similar” vs. “double”	0.001	0.0028	0.34
Image Condition “new” vs. “double”	0.0015	0.0028	0.54
Session 2 vs. 1 : Image Condition “similar” vs. “double”	0.0026	0.005	0.51
Session 2 vs. 1 : Image Condition “new” vs. “double”	-0.0037	0.0035	-1.05

Table 3: Fixed effects results in the mixed linear model of saccade amplitude in Session 2, and Image Conditions “similar”, and “new” compared with Session 1 and Image Condition “double”.

did not reveal any obvious deviations. As shown in Table 3, none of the factors had a significant effect on saccade amplitude.

The model for the third session included a fixed effect of Image Condition and random intercepts for Image and Subject, as well as a by-Subject random slope of Image Condition.

```
Amplitude ~ ImageCondition +
   (1+ImageCondition|Subject) + (1|Image)
```

There were no significant effects (see Table 4).

Similarly, investigating saccade amplitudes over the number of the fixation in a trial also did not reveal significant systematic variation (Figure 14). As in the analysis of fixation duration, we considered whether a fatigue effect of the trial number in the session may have been responsible for significant effects of presentation number on saccade amplitude in other studies. We were unable to identify such an effect (Figure 15). As before, the first trials are probably different mostly due to the new surroundings and the task.

Fixed effects			
	Estimate	Std. Error	t value
Intercept	1.0843	0.0019	559.22
Image Condition “similar” vs. “twice”	-0.0008	0.0006	-1.47
Image Condition “once” vs. “twice”	-0.0013	0.0007	-1.93
Image Condition “new” vs. “twice”	-0.0005	0.0008	-0.6

Table 4: Fixed effects results in the mixed linear model of saccade amplitude in Session 3, comparing the Image Conditions “similar”, “once” and “new” with “twice”.

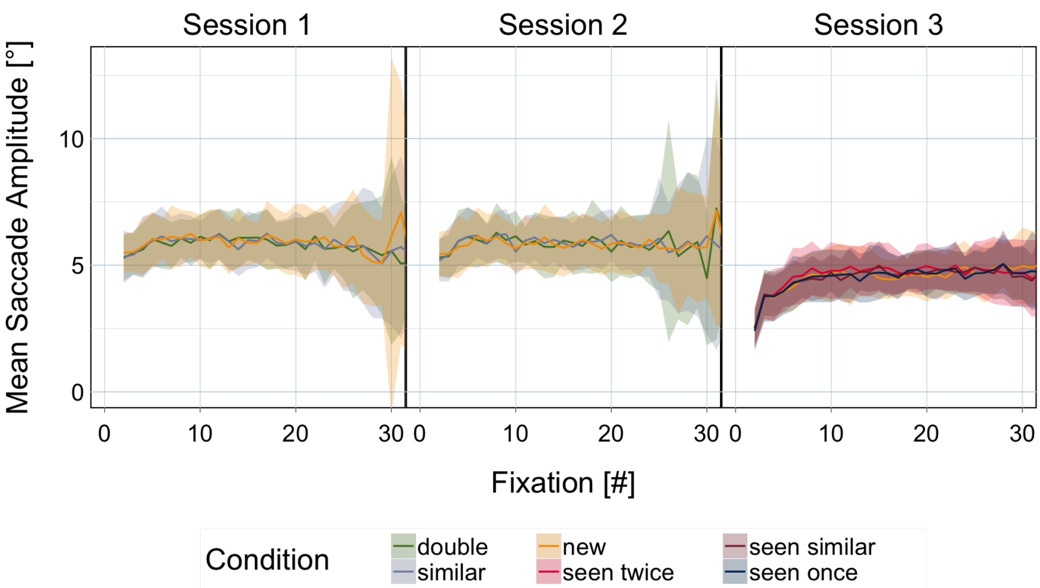

Figure 14: Saccade amplitude over the fixation number in the trial. The ribbons represent standard deviations, computed using the Cosineau method as stated in Section 3.6.

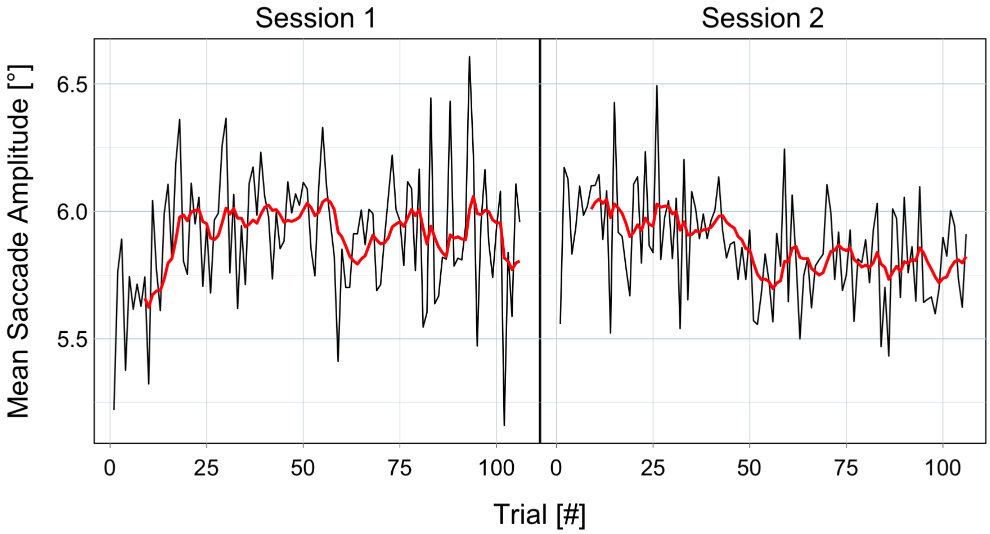

Figure 15: Saccade amplitude over ordinal trial number. The red line is a smoothing of the data using a causal kernel rolling weighted mean.

4.3 Distance to Center

Due to the central fixation bias, the measure of distance of the fixation position to the image center is not normally distributed. Rather, the distribution looks bimodal, with one peak for the first part of each trial, where there are many fixations close to the center of the image and another for later fixations which are less affected by the central fixation bias. A bimodal distribution is problematic for the linear mixed effects analysis, which assumes normally distributed residuals. For the analysis of the distance to center measurement we, therefore, used the average distance to center for each trial as a single data point.

Figure 16 shows the effect of Session and Image Condition on the mean distance to center in a trial.

The mixed linear model for the mean distance to center included fixed effects for Session, Image Condition and their interaction. The random effects structure contained a random effect of Subject and Image, and a by-Subject random slope of Session.

```
meanDTC ~ Session * ImageCondition +
    (1+Session|Subject) + (1|Image)
```

The mean distance to center measurements were transformed by raising them to the power of 1.23, in order to meet the homoscdasticity and normality assumptions. As Table 5 shows, only the comparison of Session 2 vs Session 1 was significant, i.e. participants fixated closer to the center during the second session, independently of the Image Condition.

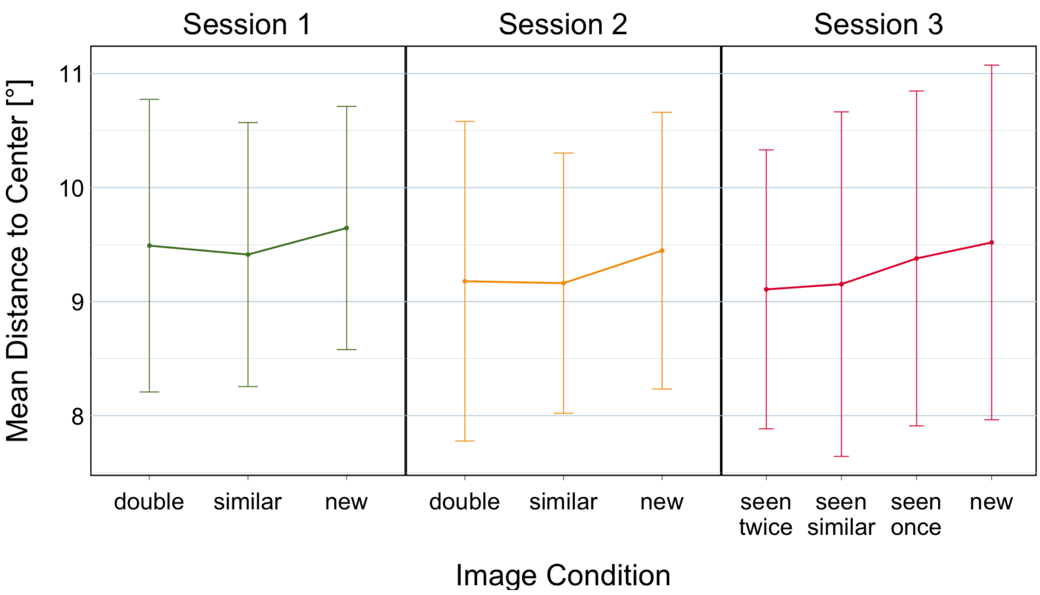

Figure 16: Mean distance to center for each trial over Session and Image Condition. The error bars represent 95% confidence intervals, computed as stated in Section 3.6.

Fixed effects			
	Estimate	Std. Error	t value
Intercept	16.1086	0.4233	38.05
Session 2 vs. 1	-0.6996	0.2242	-3.12
Image Condition “similar” vs. “double”	-0.1811	0.1439	-1.26
Image Condition “new” vs. “double”	0.2267	0.1446	1.57
Session 2 vs. 1 : Image Condition “similar” vs. “double”	0.2629	0.3358	0.78
Session 2 vs. 1 : Image Condition “new” vs. “double”	0.1772	0.2025	0.87

Table 5: Fixed effects results in the mixed linear model of the mean distance to center for each trial in Session 2, and Image Conditions “similar”, and “new” compared with Session 1 and Image Condition “double”.

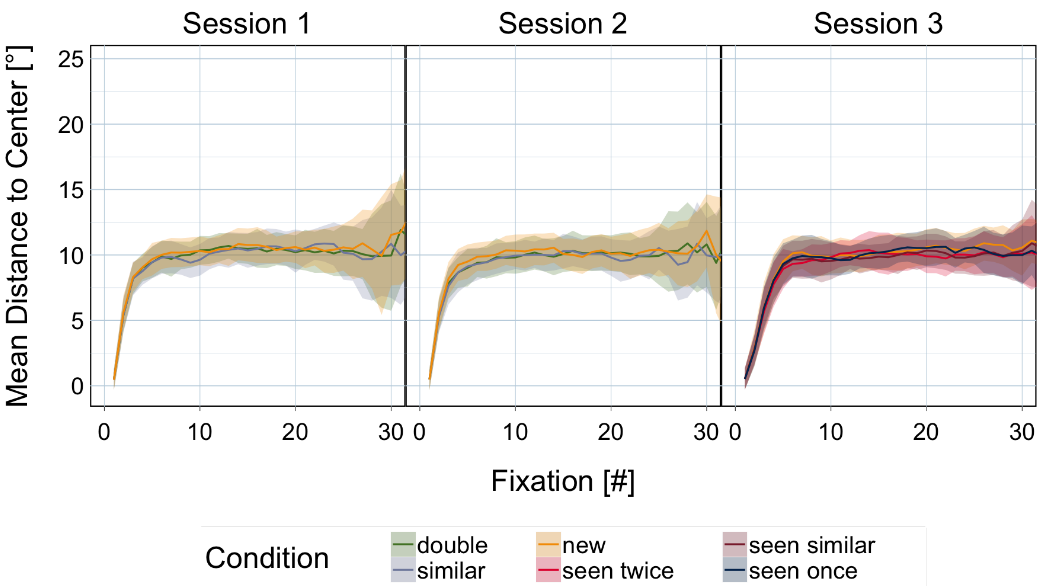

Figure 17: Distance to center over the fixation number in the trial. The ribbons represent standard deviations, computed using the Cosineau method as stated in Section 3.6.

Fixed effects			
	Estimate	Std. Error	t value
Intercept	53.8295	1.5	35.89
Image Condition “similar” vs. “twice”	0.548	0.678	0.81
Image Condition “once” vs. “twice”	3.1296	0.6817	4.59
Image Condition “new” vs. “twice”	4.1088	0.6784	6.06

Table 6: Fixed effects results in the mixed linear model of mean distance to center for each trial in Session 3, comparing the Image Conditions “similar”, “once” and “new” with “twice”.

In the linear mixed effects analysis for the mean distance to center in Session 3, we entered Image Condition as a fixed effect and random effects of Subject and Image. The measurements were raised to the power of 1.79 to meet the homoscedasticity and normality assumptions.

```
meanDTC ~ Image Condition + (1|Subject) + (1|Image)
```

The effects of Image Condition “once” vs. “twice” and Image Condition “new” vs. “twice” were significant: the distance to center was larger when the images were unfamiliar or had only been seen once before than when they had been presented in both previous sessions. It is important to note, however, that while the effect is significant, the total difference between the values is very small (less than 0.5 °).

As expected from the central fixation bias, the plot of the distance to center for each fixation in the trial (Figure 17) shows that the first fixations in the image are closer to the center than the later fixations. Apart from the initial central fixation bias, the plot does not reveal further systematic effects of the conditions.

4.4 Conformity

With the conformity measure, as described in Section 3.6, we measured how image familiarity influences whether people fixate “salient” areas as defined by the empirical fixation density of that image. As the images in the "similar" category in Session 2 were never seen in other conditions there is no empirical density data to evaluate the gaze paths on. In the mixed effects analysis of Session 1 and 2 for the conformity measurement we therefore only have two Image Conditions: "double" and "new".

Figure 18 shows that the mean conformity in Session 3 is considerably lower than in Sessions 1 and 2. As participants were restricted in their visual field, they were unable to identify the locations that were salient in the regular viewing sessions. The values in Session 1 serve as a baseline for how similar the fixated regions of an observation are to the empirical density under ideal viewing conditions when the image is unfamiliar.

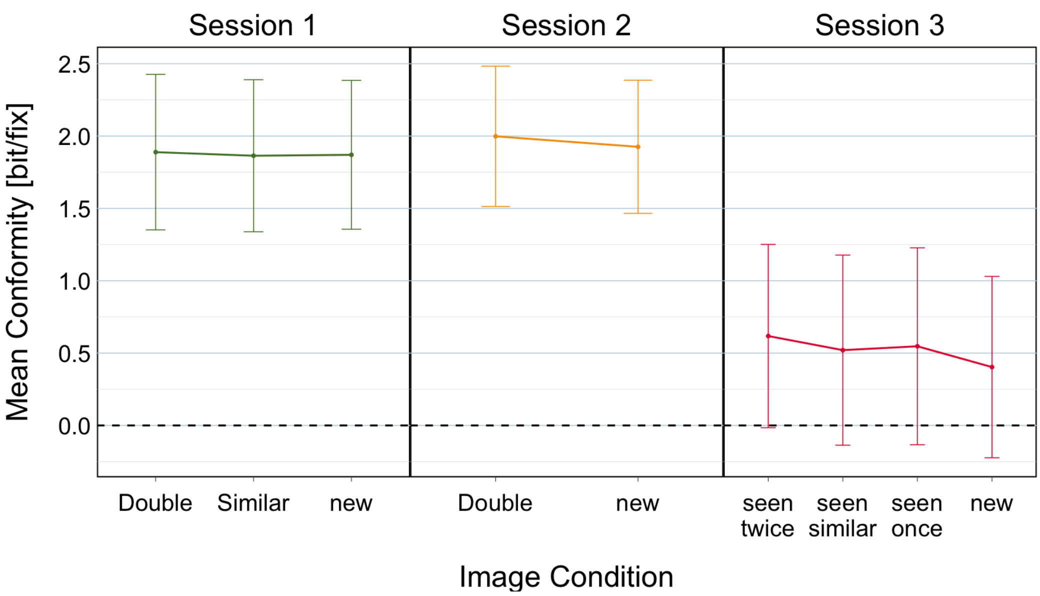

Figure 18: Conformity over Session and Image Condition, showing how much better the empirical density describes the fixation data than a uniform distribution. The error bars represent 95% confidence intervals, computed as stated in Section 3.6.

As before, we used Image Condition, Session and their interaction as main effects and random intercepts of Subject and Image. We also added

Fixed effects			
	Estimate	Std. Error	t value
Intercept	-2.2287	0.0047	-471.81
Session 2 vs. 1	0.007	0.0031	2.29
Image Condition "new" vs. "double"	-0.0002	0.0025	-0.09
Session 2 vs. 1 : Image Condition "new" vs. "double"	-0.0052	0.0026	-2.01

Table 7: Fixed effects results in the mixed linear model of the conformity in Session 2 compared with Session 1, and Image Condition "new" compared with "double".

random by-Subject slopes of Session and Image Condition and the maximum by-Image random slopes.

```
Conformity ~ Session * ImageCondition +
    (1+Session + ImageCondition|Subject) +
    (1+Session * ImageCondition|Image)
```

The Box-Cox transform which we used for the other measures to make the data conform to the homoscdasticity and normality assumptions takes strictly positive input. As the conformity values can be negative, we instead used the Yeo-Johnson transform for this model. However, even with using the Yeo-Johnson transform, the residual distributions of the data did not become truly normally distributed. There can be extremely low values of conformity, causing a near-normal distribution with a long one-sided tail. The following results were obtained from these data. However, we did repeat the analysis under exclusion of the outliers and the results were not significantly different.

As Table 7 shows, we found a significant effect of Session 1 vs Session 2, i.e. conformity values in Session 2 were higher. There was also a significant interaction of Session 2 vs. 1 and Image Condition "new" vs. "double". Conformity values were, therefore, higher in Session 2 when they were in the "double" condition than in the "new" condition.

Fixed effects			
	Estimate	Std. Error	t value
Intercept	-2.309	0.0044	-524.34
Image Condition “similar” vs. “twice”	-0.0069	0.0029	-2.4
Image Condition “once” vs. “twice”	-0.0051	0.0031	-1.64
Image Condition “new” vs. “twice”	-0.0145	0.004	-3.66

Table 8: Fixed effects results in the mixed linear model of conformity in Session 3, comparing the Image Conditions “similar”, “once” and “new” with “twice”.

The linear mixed model for Session 3 included a fixed effect of Image Condition, random intercepts for Subject and Image, and a by-Subject slope of Image Condition.

```
Conformity ~ ImageCondition +
    (1+ ImageCondition|Subject) + (1|Image)
```

The model for conformity in Session 3 should be considered under the same caveats of the homoscedasticity and normality assumptions as the model for Session 1 and 2. Table 8 shows that there was a main effect of Image Condition “similar” vs. “twice” and a main effect of Image Condition “new” vs. “twice”. In other words, in Session 3, conformity with the empirical density was lower when the image had never been seen or when it had been seen in Session 1 and a semantically similar image had been seen in Session 2, compared to when it had been seen twice.

Examining how conformity evolves over the fixations in the trial reveals that in all sessions and conditions the gaze path becomes more dissimilar from the empirical density over time (Figure 19). The initial peak is likely due to the central fixation bias, as all participants started their inspection in the center. In Session 3 conformity descends all the way down to 0, meaning that the fixations can not be predicted better by the empirical density than by a uniform distribution. However, as seen in the analysis, the “seen twice” condition stays slightly higher, especially during the later fixations. A higher conformity value indicates that participants were better able to use information in the image as they could remember information about the image.

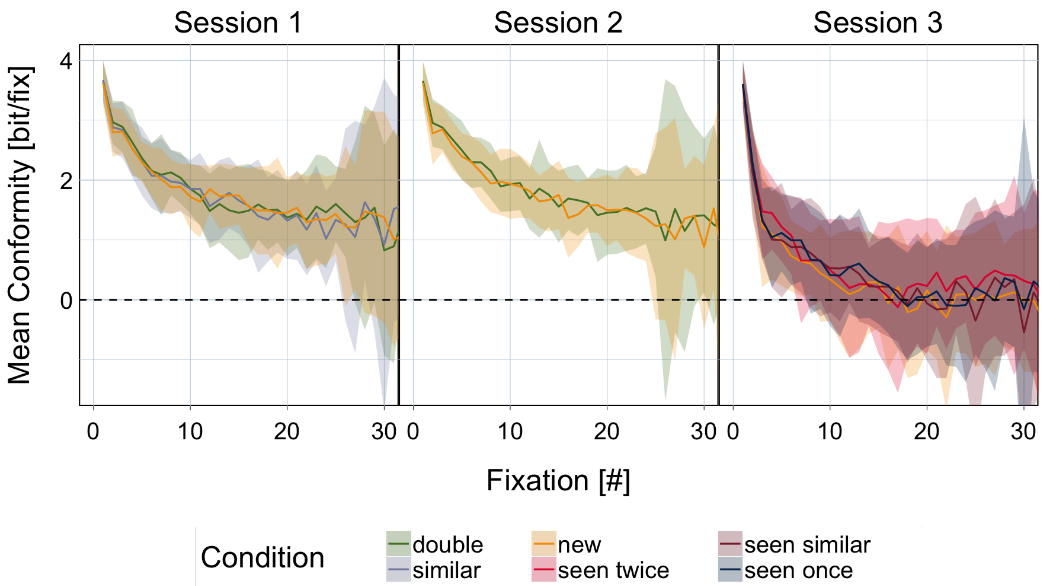

Figure 19: Conformity over the fixation number in the trial, showing how much better the empirical density describes the fixation data than a uniform distribution. The ribbons represent standard deviations, computed using the Cosineau method as stated in Section 3.6.

4.5 Familiarity

In the recall part of Session 1 and 2 and during Session 3 we recorded participants’ answers to the question “have you seen this image before?”. With the analysis of the given answers we confirm that the familiarity manipulation worked. Participants showed extraordinarily high levels of proficiency in recognizing the images (Figure 20). In the recalls from Session 1 and 2 they had to remember images that had been shown to them during the same session. In Session 3 they indicated for each image whether it was familiar.

Even though information was reduced by the moving window, participants were able to correctly classify all categories of images at higher than chance level. Even images from the "seen once" condition, which was only shown once in the first session were mostly correctly recognized.

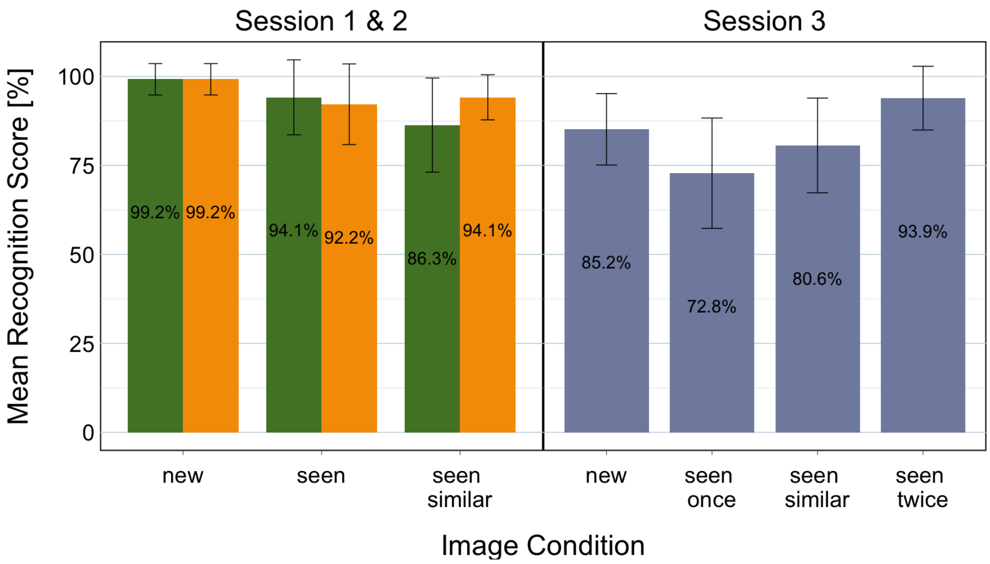

Figure 20: Average participant scores in the recall task. In Session 1 and 2 participants were asked to indicate whether images had been shown in the same session. The green are results from the first session, yellow are from the second session. In Session 3 participants viewed the images with a gaze contingent moving window and indicated whether the image was familiar after each trial. The error bars represent standard deviations from the mean between the participants.

To summarize, we found the following significant effects for the comparison of Image Conditions over Session 1 and 2:

- main effect of Image Conditions "new" compared with "double" on fixation duration
- main effect of Session on distance to center
- main effect of Session on conformity

- interaction effect on conformity of Session and Image Conditions “new” compared with “double”.

For the gaze contingent moving window paradigm in Session 3 we found significant effects of

- Image Conditions “once” and “new” compared with “twice” on distance to center
- Image Conditions “similar” and “new” compared with “twice” on conformity.

These effects and also the absence of others will be discussed in detail in the following section (Section 5).

5 Discussion

The aim of the present experiment was to explore the effects of visual long-term memory on eye movement. An effect of image familiarity can be found when images are repeatedly presented within one experimental session. In the present experiment we explored whether the familiarity effect can be extended to longer time scales with images that are presented several days apart. For this purpose, we compared participants' eye movements in response to unfamiliar, familiar, and semantically familiar images.

Replication of the familiarity effect. For the measures fixation duration, saccade amplitude, and distance to center we found no significant interaction effects between the image familiarity conditions and Session 1 and 2. That is, there was no systematic difference in these measures whether the image had been seen before or whether it was new or similar to a seen image. The only effects we found were a main effect of Session for the distance to center measure and a main effect of Image Condition on fixation duration. As both sessions were conducted in identical setups and the images were randomly assigned to the conditions, these effects are unlikely to be meaningful. Inspecting these measures over the ordinal fixation number in the trial also did not reveal any differential trends.

The lack of significant effects of image familiarity is unexpected and stands in direct contrast to the results of Kaspar and König (2011b, 2011a) and Trukenbrod et al. (2017). Based on these studies we expected to find an increase in fixation duration and a decrease in saccade amplitude when the image was familiar compared to when it was new. The absence of these effects in the present experiments indicates that either the image material was inadequate or that the familiarity effect found in previous studies is very short-lived. Further testing is required to determine the cause for this inconsistency with previous literature (see Section 5.1).

However, the latter explanation would cast doubt on a direct link between familiarity and eye movement behavior. As we showed, participants were well familiar with the images. They were able to identify which images had been previously presented and which had not, even under non-ideal viewing conditions. Despite their familiarity with the images, participants' eye movement measures did not reveal systematic differences between familiar and unfamiliar images.

We considered the possibility that the effects found in other studies may have been confounded with an effect of trial number in a session. The second presentation must occur after the first presentation and eye movements may change over the course of a session, e.g. due to tiredness. Although at first inspection we did not find an effect of trial fatigue with the measures of fixation duration and saccade length, some manner of fatigue effect may still be relevant.

Alternatively, we hypothesize that short-term familiarity effects may be caused by remaining motor activation from the first presentation. When the time between presentations is not sufficiently long, activation has not fully dissipated, causing eye movement behavior to change systematically. When several days have passed between presentations, however, the reaction to a familiar image is only mediated by long-term memory and not by residual activation.

At this point we cannot conclusively answer Research Question 1 and 4, but there is reason to believe that the familiarity effect as it was described previously, does not persist over long time spans.

Conformity. The conformity of a gaze path with the empirical density of fixations is a measure of whether the same regions are considered "salient" in an individual viewing as the empirical fixation density indicates. We expected conformity to vary with image familiarity as different areas are likely to be interesting or different strategies to be relevant in a familiar compared to an unfamiliar image.

We found a small but significant effect of image familiarity on conformity. Conformity values were higher in familiar images in the second session than in the first, meaning the empirical density was a better predictor of the gaze path. This result implies that participants were using their top-down knowledge to adapt viewing behavior. As movement measures like fixation duration and saccade amplitude were not affected, we hypothesize that the influence of long-term image memory is limited to the fixation locations, and affects movement measures only to a lesser extent.

Semantic similarity. Semantically similar images were included in order to gain information about how images are remembered: on a semantic or on a specific level. If the familiarity effect applies to semantically similar images as it does to familiar images, this would indicate mostly semantic information is remembered.

As we were unable to reproduce the familiarity effect, however, Research Question 3, that is whether the familiarity effect relies on semantics or a specific image, can not be conclusively answered by this thesis.

In the recall, participants were slightly worse at correctly rejecting images that were semantically similar to seen images, especially in Session 1. In Session 2 they improved their performance, indicating that they learned to pay more attention to detail once they knew the difficulty of the recall task. Knowledge about the task requirements seems to exert an influence on how well images can be memorized.

Session 3. The conditions and setup in Session 3 differed significantly from the previous sessions. Here we split the conditions into (a) seen twice, (b) seen once in Session 1 and seen a similar image in Session 2, (c) seen once, in session 1, (d) new. Thus, these conditions form a scale of how familiar the participant should be with the image. The gaze-contingent moving window manipulation was intended to produce eye movements that are mostly memory (top-down) driven. Assuming that the manipulation worked as intended, the influence of familiarity on eye movement should be more evident in this session.

Nonetheless, there were no effects of familiarity on the movement measures of eye movement, i.e. saccade amplitude and fixation duration. There were, however, significant effects on the average distance to center in a trial and on the conformity of the gaze path with the empirical density.

In the absence of peripheral information, participants do rely on information from memory to guide their eye movements in terms of fixation locations. However, effects of memory did not influence *how* the movements were made.

For the distance to center measurement the conditions differed mostly at the very beginning and at the end of the trial. Participants initially stayed closer to the center and also returned to be closer to the center later in the trial. In the case of conformity with the empirical density, the difference between twice seen images and new images is, again, most evident at the very beginning and at the end of the trial.

We can speculate that the difference at the beginning is linked to initial recognition, causing participants to stay in the center longer to identify the familiar aspects of the current area. Due to the central fixation bias, conformity values are higher in the center of the image. The later difference might be caused by lack of further exploration in familiar images, leading participants to return to the center, where conformity is high. As the task in Session 3 was to recall the images, once participants were certain of their answer, there was no motivation to continue exploring.

It is also important to be aware that the differences between the conditions, while significant, are quite small.

5.1 Limitations and Future Research

The differences between the image familiarity conditions were unexpectedly small in this experiment. The effects on movement measures that have been previously reported were absent in these data and even the effects on location measures were not large.

One important caveat to the experimental setup was that the time between the sessions was vastly different from one participant to another. While the majority completed all sessions within 5 or 6 days, some participants took over 10 days. In order to be able to interpret the results, this inconsistency should be remedied.

In a follow-up study we will conduct the same experiment but with the first two sessions on the same day and the third session on the next. If the familiarity effect is stronger in the follow-up experiment, it would lend strong evidence to the idea that there is a time component in the influence of image familiarity on eye movement. However, as participants do still remember the images well even after many days, it would also cast doubt on familiarity as the sole cause of the observed effect.

If the effect is also absent in the described follow-up study, a third study may be required to exactly replicate the previous studies using our image material and setup, to confirm the null result.

Another point of limitation is that Kaspar and König (2011b, 2011a) found the strongest effects on fixation duration and saccade amplitude in images that had been repeated multiple times. In the study described here repeated images are only shown up to three times. While the effect can be seen even in the second presentation and has also been replicated for the second presentation by Trukenbrod et al. (2017), it is possible that the repeated exposure, possibly in addition to the shorter time frame, were crucial to the effect.

6 Conclusion

This thesis explores the effect of visual long-term memory on eye movement. Previous studies have found that familiarity with an image affects viewing behavior measures such as fixation duration and saccade amplitude. However, the literature on the the effect of image familiarity on eye movement was limited to image repetition within one experimental session. Human visual long-term memory has a high capacity and is extraordinarily long lasting, far exceeding one experimental session. In this study, images are repeated over three sessions, conducted over the course of multiple days.

With the longer time between the presentations, we were unable to replicate the familiarity effect on fixation duration and saccade amplitude. We did find evidence of small effects of image familiarity on conformity of individual gaze paths with the empirical density of fixations on an image, and on the average distance to the image center within a trial. We speculate that image familiarity on a longer time scale may still influence the locations of fixations, but not the movements themselves. More generally, however, our results indicate that scene exploration is primarily driven by the current visual input and only weakly influenced by visual long-term memory from previous days.

The absence of the previously found familiarity effect is unexpected, especially because participants did demonstrate very good memory of the images. A follow up study of the same design but with shorter times between the sessions will need to be conducted so as to appropriately interpret the current results.

References

Aloimonos, Y., & Rosenfeld, A. (1991). Computer vision. *Science*, *253*(5025), 1249–1254.

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

Arons, B. (1992). A review of the cocktail party effect. *Journal Of The American Voice I/o Society*, *12*(7), 35–50.

Baayen, H., Bates, D., Kliegl, R., & Vasishth, S. (2015). Repsychling: Data sets from psychology and linguistics experiments [Computer software manual]. (R package version 0.0.4)

Baddeley, A., Rubak, E., & Turner, R. (2015). *Spatial point patterns: Methodology and applications with R.* London: Chapman and Hall/CRC Press. Retrieved from `http://www.crcpress.com/Spatial-Point-Patterns-Methodology-and-Applications-with-R/Baddeley-Rubak-Turner/9781482210200/`

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

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *ArXiv*. Retrieved from `https://arxiv.org/abs/1506.04967`

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

Bear, M. F., Connors, B. W., & Paradiso, M. A. (Eds.). (2007). *Neuroscience* (2nd ed.). Baltimore, MD: Lippincott Williams & Wilkins.

Becker, M. W., Pashler, H., & Lubin, J. (2007). Object-intrinsic oddities draw early saccades. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(1), 20–30.

Borji, A., & Itti, L. (2014). Defending yarbus: Eye movements reveal observers' task. *Journal of vision*, *14*(3), 29.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, *11*(5), 4.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*(38), 14325–14329.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.

Broadbent, D. E. (1954). The role of auditory localization in attention and memory span. *Journal of Experimental Psychology*, *47*(3), 191–196.

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

Bylinskii, Z., Judd, T., Borji, A., Itti, L., Durand, F., Oliva, A., & Torralba, A. (n.d.). *Mit saliency benchmark.*

Castelhano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 753–763.

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.

Castet, E., & Masson, G. S. (2000). Motion perception during saccadic eye movements. *Nature Neuroscience*, *3*(2), 177–183.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, *25*(5), 975–979.

Churchland, P. S., Ramachandran, V. S., & Sejnowski, T. J. (1994). A critique of pure vision. In C. Koch & J. L. Davis (Eds.), *Large-scale neuronal theories of the brain* (pp. 23–60). Cambridge, MS: MIT Press.

Collewijn, H., Erkelens, C. J., & Steinman, R. M. (1988). Binocular co-ordination of human horizontal saccadic eye movements. *The Journal of Physiology*, *404*(1), 157–182.

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.

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

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

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

Engbert, R., Sinn, P., Mergenthaler, K., & Trukenbrod, H. (2015). Microsaccade toolbox for r. Retrieved from `http://read.psych.uni-potsdam.de/attachments/article/140/MS_Toolbox_R.zip`

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

Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing* (No. 37). Oxford, UK: Oxford University Press.

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. (2012). Modelling the influence of central and peripheral information on saccade biases in gaze-contingent scene viewing. *Visual Cognition*, *20*(4-5), 546–579.

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.

Godwin, H. J., Reichle, E. D., & Menneer, T. (2014). Coarse-to-fine eye movement behavior during visual search. *Psychonomic Bulletin & Review*, *21*(5), 1244–1249.

Greene, M. R., Liu, T., & Wolfe, J. M. (2012). Reconsidering yarbus: A failure to predict observers' task from eye movement patterns. *Vision Research*, *62*, 1–8.

Harris, C. R., Kaplan, R., & Pashler, H. (2008). Alarming events in the corner of your eye: Do they trigger early saccades? *Social Science Research Network*. Retrieved from `https://ssrn.com/abstract=2542346`

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

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

Hoffman, J. E. (1998). Visual attention and eye movements. *Attention*, *31*, 119–153.

Hollingworth, A. (2006). Visual memory for natural scenes: Evidence from change detection and visual search. *Visual Cognition*, *14*(4-8), 781–807.

Horváth, G. (2017). *Rawtherapee 5.3.* Retrieved from `http://rawtherapee.com/`

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 *2009 IEEE 12th international conference on computer vision.* IEEE.

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

Kaspar, K., & König, P. (2011b). Viewing behavior and the impact of low-level image properties across repeated presentations of complex scenes. *Journal of Vision*, *11*(13), 26.

Klein, R. M., & MacInnes, W. J. (1999). Inhibition of return is a foraging facilitator in visual search. *Psychological Science*, *10*(4), 346–352.

Kowler, E. (2011). Eye movements: The past 25years. *Vision Research*, *51*(13), 1457–1483.

Kümmerer, M., Theis, L., & Bethge, M. (2014). Deep gaze i: Boosting saliency prediction with feature maps trained on imagenet. *Arxiv*. Retrieved from `https://arxiv.org/abs/1411.1045`

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

Kümmerer, M., Wallis, T. S. A., Gatys, L. A., & Bethge, M. (2017, Oct). Understanding low- and high-level contributions to fixation prediction. In *The ieee international conference on computer vision (iccv).*

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

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.

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

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* New York, NY: Freeman.

Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*, *5*(3), 229–240.

Martinez-Conde, S., Macknik, S. L., Troncoso, X. G., & Dyar, T. A. (2006). Microsaccades counteract visual fading during fixation. *Neuron*, *49*(2), 297–305.

MathWorks Inc. (2017). *Matlab 2017b (version 9.3.0).* Natick, Massachusetts.

Mathôt, S., Melmi, J.-B., & Castet, E. (2015, aug). Intrasaccadic perception triggers pupillary constriction. *PeerJ*, *3*, e1150.

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

Matthis, J. S., Yates, J. L., & Hayhoe, M. M. (2018). Gaze and the control of foot placement when walking in natural terrain. *Current Biology*, *28*(8), 1224–1233.

Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. *Journal of Vision*, *6*(1), 2.

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.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*(2), 61–64.

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

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

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *46*(3), 461–488.

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.

Parkhurst, D., & Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision*, *16*(2), 125–154.

Pelz, J. B., & Canosa, R. (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research*, *41*(25-26), 3587–3596.

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.

Posner, M., & Cohen, Y. (1984). Components of visual orienting. *Attention and performance X: Control of language processes*, *32*, 531–556.

R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

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

Rayner, K. (2009). The 35th sir frederick bartlett lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506.

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

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.

Schwetlick, L., Rothkegel, L. O. M., Trukenbrod, H. A., & Engbert, R. (2017). Central fixation bias: The role of sudden image onset and early gist extraction. In *European conference on visual perception 2017.*

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.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, *6*(1), 156–163.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, *1*(7), 261–267.

Simons, D. J., & Rensink, R. A. (2005). Change blindness: past, present, and future. *Trends in Cognitive Sciences*, *9*(1), 16–20.

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

Standing, L. (1973). Learning 10000 pictures. *Quarterly Journal of Experimental Psychology*, *25*(2), 207–222.

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

Strasburger, H., Rentschler, I., & Juttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision*, *11*(5), 13.

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

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.

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.

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*. Retrieved from https://arxiv.org/abs/1704.01761v1

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.

Wolfe, J. M., Kluender, K. R., Levi, D. M., Bartoshuk, L. M., Herz, R. S., Klatzky, R. L., … Merfeld, D. M. (2006). *Sensation & perception* (2nd ed.). Sunderland, MA: Sinauer Associates.

Wood, N., & Cowan, N. (1995). The cocktail party phenomenon revisited: How frequent are attention shifts to one's name in an irrelevant auditory channel? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 255–260.

Yantis, S., & Jonides, J. (1996). Attentional capture by abrupt onsets: New perceptual objects or visual masking? *Journal of Experimental Psychology: Human Perception and Performance*, *22*(6), 1505–1513.

Yarbus, A. L. (1967). *Eye movements and vision.* New York, NY: Plenum Press.

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig verfasst, sowie keine anderen Quellen und Hilfsmittel als die angegebenen benutzt habe.

Potsdam, 27. Juni 2018

Lisa Schwetlick