Temporal evolution of the central fixation bias in scene viewing

Lars O. M. Rothkegel
University of Potsdam, Potsdam, Germany

Hans A. Trukenbrod
University of Potsdam, Potsdam, Germany

Heiko H. Schütt
University of Potsdam, Potsdam, Germany
Eberhard Karls University, Tübingen, Germany

Felix A. Wichmann
Eberhard Karls University, Tübingen, Germany
Max Planck Institute for Intelligent Systems, Tübingen, Germany
Bernstein Center for Computational Neuroscience, Tübingen, Germany

Ralf Engbert
University of Potsdam, Potsdam, Germany

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

Introduction

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

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

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

To investigate the principles underlying the reduced CFB, we designed and analyzed four experiments, in which observers started exploration from different positions within an image and were required to maintain fixation for various time intervals after image onset (pretrial fixation time). Our study used the images investigated in the most frequently cited paper on the central fixation bias (Tatler, 2007) to exclude any influence of the images on the reduction of the CFB.

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

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

### General methods

#### Stimuli

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

#### Participants

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

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

Data analysis

Data preprocessing and saccade detection

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

Distance to center over time

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

\[
DTC_T = \frac{1}{m \cdot n} \sum_{j=1}^{n} \sum_{k=1}^{m} ||x_{jk}(t) - x_{\text{center}}||, \quad (1)
\]

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

Influence of the initial fixation on the second fixation

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

Density maps of eye positions over time

To visualize the temporal evolution of eye positions in our experiments, we computed movies of two-dimensional density maps for the different pretrial fixation times and each eye position of the time series recorded for each experiment. Based on a kernel density estimation via diffusion (Botev, Grotowski, & Kroese, 2010), we estimated density maps for the first 2 s (after removal of the fixation cross) in each experiment. These movies are available as supplementary material.

Experiment 1

Methods

Participants

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

**Procedure**

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

**Results**

**Distance to center over time**

In Experiment 1, the DTC initially decreased for all conditions (i.e., the CFB increased; see Figure 2a). There was a pronounced effect that mean fixation
positions tended to be closer to the image center when participants were allowed to explore an image immediately after image onset, i.e., with a pretrial fixation time of 0 ms (black curves in Figure 2a). Surprisingly, for the first four participants (Block 1) of this group the effect was visible throughout the whole observation time of 5 s. A second group of participants in the 0 ms condition (Block 2) did not replicate the stronger CFB through the whole observation time. In addition, there was a gradual reduction of the CFB for pretrial fixation times from 125 ms to 250 ms (red and green curve). DTC for pretrial fixation times of 250 and 500 ms hardly differed (green vs. blue curve). The minimum for the pretrial fixation time of 1000 ms occurred later in time because of disproportionately long saccade latencies of the first saccade after a forced fixation on the fixation cross of 1000 ms (cyan curve).

**Distance to center on the second fixation**

Figure 2b shows the influence of initial saccade latency on the mean DTC of the second fixation for the five pretrial fixation times. Each bin represents a quintile of the distribution of saccade latencies in each condition. A clear relation between DTC and latency of the initial saccade is visible for the pretrial fixation times of 0 ms and 125 ms. Overall, short saccade latencies led to a small average DTC (i.e., a strong initial CFB) whereas long latencies led to a larger average DTC (i.e., a less pronounced initial CFB).

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

**Discussion**

Experiment 1 led to a reduction of the CFB on the initial saccade target for all pretrial fixation times of 125 ms and more during scene perception from extreme starting positions (Figure 2a). A pretrial fixation time of 125 ms produced an intermediate CFB, whereas longer pretrial fixation times produced an asymptotic behavior. With a pretrial fixation time of 0 ms the DTC was smaller throughout almost the whole observation time of 5 s for the first group of participants. However, this effect was not replicated in a retest with 20 new participants. The early effect of the CFB did not differ in the two groups of participants. The CFB of the second fixation did strongly depend on the latency of the initial saccade (Figure 2b). Thus, the early differences between pretrial fixation times in Figure 2a are driven by differences in the distribution of initial saccade latencies.

These results replicated our earlier findings of a reduced CFB during scene perception by introducing a non-zero pretrial fixation time (Rothkegel et al., 2016). A delay of 125 ms was sufficient to achieve a considerable reduction and after a delay of 250 ms the minima of DTC curves only differed marginally. In addition, our results suggest that the most important mediating factor of the CFB was the latency of the first saccadic response. Saccades with brief saccade latencies were on average directed more strongly toward the center than saccades with long saccade latencies.
Experiment 2

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

Methods

Participants

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

Procedure

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

Results

Distance to center over time

In Experiment 2, where the starting positions were located on a ring around the image center, the eyes moved initially even further towards the image center in the 0 ms pretrial condition (black curve in Figure 3a was the only curve with a pronounced negative slope in the beginning). A difference in DTC was visible until about 600 ms after offset of the fixation marker. Later during the trial, the curves converged for all pretrial conditions and reached a stable DTC for the rest of the trial. Qualitatively, we also observed a small initial difference in DTC between short pretrial fixation times of 125 ms and 250 ms and pretrial fixation times of 500 ms and 1000 ms.

Distance to center of the second fixation

As in Experiment 1, we found a strong influence of the latency of the first saccade on the DTC of the second fixation for small pretrial fixation times (Figure 3b). The results of the linear mixed model in Figure 3. Experiment 2. (a) Mean distance to center over time [DTC(t)] for the five different pretrial fixation times with starting positions on a donut-shaped ring around the image center. Confidence intervals indicate SE as described by Cousineau (2005). (b) Mean distance to center of the second fixation as a result of initial saccade latency and pretrial fixation time. Bins represent quintiles of the saccade latency distribution. Error bars are the SEM.

Experiment 2 (Table 2) were similar to Experiment 1. The most important results are the significantly lower DTC of the 0 ms pretrial fixation time compared with the average and the significant increase in DTC for
higher saccade latencies. As in Experiment 1 an interaction between saccade latency and pretrial fixation time is visible. This is especially true for the 0 ms condition, where the influence of saccade latency significantly increases compared with the average influence. In Experiment 2 the only significant decrease in saccade latency influence is visible for a pretrial fixation time of 250 ms. Overall direction of the influence (increasing influence of saccade latency for pre-trial fixation times of 0 ms and 125 ms vs. decreasing influence for pretrial fixation times of 250 ms and 500 ms) is the same as in Experiment 1.

Discussion

If the starting position was close to the image center all pretrial fixation times of 125 ms or longer (Figure 3a) led to a reduction of the CFB on early fixations. After around 600 ms this influence disappeared. Furthermore, a clear relation between latency of the first saccade and the CFB of the second fixation was visible (Figure 3b). Thus, the results replicated our observations from Experiment 1 and demonstrated that a reduced CFB was not exclusively generated by the extreme starting positions used in Experiment 1.

Experiment 3

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

Methods
Participants

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

Procedure

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

Results

Distance to center over time

As in Experiment 1 and 2 the eyes initially moved towards the center for all pretrial fixation times (Figure 4a). The difference between pretrial conditions was not as clearly visible as in previous experiments. Even the difference between the 0- and the 125-ms condition was relatively small. The smaller difference was probably due to the blocked design where pretrial fixation times changed after 20 trials during the experiment for each participant. Nonetheless, curves with a pretrial fixation time smaller than or equal to 50 ms had smaller minima than the ones with pretrial fixation times larger than 50 ms (see inset in Figure 4a).
The influence of the first saccade latency on the distance to center of the second fixation is clearly visible in Figure 4b. The influence seemed even clearer than in previous experiments. However, the range of the distance to center values was larger in this experiment as a result of the increased magnitude of the image in visual degree. Saccade latencies were more homogeneous in Experiment 3. The difference of mean saccade latencies (pretrial fixation time + saccade latency after removal of the fixation marker) between the 0- and 125-ms condition was much smaller (57 ms) than in Experiments 1 (154 ms) and 2 (138 ms).

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

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

Discussion

Experiment 3 was conducted to investigate the minimum pretrial fixation time necessary for a reduction of the early central fixation bias. All pretrial conditions showed a similar behavior with a tendency of an early CFB as measured by the DTC. We observed the weakest DTC effect for pretrial fixation times of 125 ms (inset in Figure 4a). Pretrial fixation times equal to or smaller than 50 ms generated fixation positions closest to the image center. Differences in DTC could be explained by the influence of the first saccade latency on the selection of the second fixation location (Figure 4b). Thus, saccade latencies are the most important factor modulating the CFB. A post-hoc analysis revealed that saccade latencies were only affected in conditions with pretrial fixation times larger than 50 ms. This is in line with previous research that the shortest image preview to influence further eye movement behavior in visual search lies between 50 and 75 ms (Vö & Henderson, 2010). We conclude that a minimum pretrial fixation time of around 75 ms is needed to prolong saccade latencies in order to reduce the CFB in scene viewing.

Experiment 4

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

Methods

Participants

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

Procedure

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

Results

Distance to center over time

Contrary to the first experiments initial gaze positions could only move away from the image center with central starting positions in Experiment 4 (Figure 5a). Therefore, DTC gradually increased until it reached an asymptote. Between pretrial conditions, DTC differed with respect to the point in time, when curves started to monotonically increase (pretrial fixation times: 250 ms < 125 ms < 0 ms). Although pretrial fixation times were chosen to be equidistant, curves for 125-ms and 0-ms pretrial conditions (red and black curve) take longer to converge than curves for the 250-ms and 125-ms pretrial conditions (green and red curve; see inset Figure 5a). This demonstrated that pre-trial fixation times of 125 ms or more reduce the CFB of early fixations even during scene viewing with central starting positions.

Distance to center of the second fixation

Latencies of the first saccade were longer in this experiment than in any of the other experiments. This observation is in line with results from face perception, where the initial fixation is longer when participants start exploring a face in the center (Arizpe, Kravitz, Yovel, & Baker, 2012). Due to the increased number of long initial saccade latencies, an influence of saccade latency on the second fixation location was not as clearly visible as in the previous experiments (Figure 5b). Results of a linear mixed model for Experiment 4 partially replicated the main results from Experiments.

<table>
<thead>
<tr>
<th>Fixed effect</th>
<th>Estimate</th>
<th>SE</th>
<th>t</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>2.858</td>
<td>0.157</td>
<td>18.194***</td>
<td></td>
</tr>
<tr>
<td>0 ms</td>
<td>-0.309</td>
<td>0.093</td>
<td>-3.312***</td>
<td></td>
</tr>
<tr>
<td>125 ms</td>
<td>0.155</td>
<td>0.083</td>
<td>1.877</td>
<td></td>
</tr>
<tr>
<td>saccade latency</td>
<td>0.224</td>
<td>0.122</td>
<td>1.837</td>
<td></td>
</tr>
<tr>
<td>saccade latency × 0 ms</td>
<td>0.454</td>
<td>0.174</td>
<td>2.611**</td>
<td></td>
</tr>
<tr>
<td>saccade latency × 125 ms</td>
<td>-0.189</td>
<td>0.154</td>
<td>-1.222</td>
<td></td>
</tr>
</tbody>
</table>

Random Effects

| variance: Subjects   | 0.26 |
| variance: Images     | 0.18 |

Table 4. Output of LMM for Experiment 4. Notes: *p < 0.05, **p < 0.01, ***p < 0.001.
The DTC of the second fixation was significantly smaller for a pretrial fixation time of 0 ms. The influence of saccade latency on distance to center of the second fixation did not reach a level of significance of 95% in Experiment 4. The direction of the influence was positive and nearly reached the level of significance. The fact that saccade latency was not a significant predictor is a result of the rather long latencies and a small number of participants. By removing initial saccade latencies of higher than 1 s (which normally are very rare) saccade latency becomes a significant predictor ($p < 0.03$). The interaction between saccade latency and pretrial times showed that the influence of saccade latency on DTC was, as observed in Experiments 1 and 2, significantly larger for a pretrial fixation time of 0 ms.

Discussion

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

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

Discussion of empirical results

In four scene-viewing experiments we have shown that by delaying the initial saccade relative to the sudden image onset the early central fixation bias was
significantly reduced. Further analysis showed that the amount of early CFB is directly linked to the initial saccade latency. Figure 6 shows the influence of initial saccade latency on distance to image center for all four experiments combined. A clear increase of DTC is visible between 150 and 400 ms. Because initial saccade latencies above 400 ms do not show an influence, pretrial fixation times above 250 ms did not produce noteworthy effects. This also explains why in Experiment 4 the rather long saccade latencies were not a significant predictor for the CFB. We conclude our experiment by stating that the initial saccade latency is the dominant factor influencing the early central fixation bias in scene viewing. This leads to the assumption that the sudden image onset is involved in generating the early CFB.

Computational modeling of the central fixation bias

To test if the early CFB might result from default activation in the image center after a sudden onset that is replaced by a content driven activation over time, we simulated scanpaths generated by a computational model. For the simulations we used an extended version of the previously published SceneWalk model of saccade generation from our group (Engbert et al., 2015). Different to the original model with zero activation at the beginning of a trial, we decided to start each trial with higher activations in the center of an image than at the periphery (see Figure 7a). The influence of this central starting activation declines with increasing saccade latency and is replaced by a more content driven activation (the empirical density map of the image multiplied with a Gaussian around the starting position; see Figure 7b). This initial central activation represents the sudden image onset. We refer to this extended model as the SceneWalk StartMap model. A more detailed description of the model can be found in the Appendix.

Figure 7 shows the simulated fixations 1–4 of two trials with different pretrial conditions (0 ms vs. 1000 ms) of the SceneWalk StartMap model. The initial saccade latency in the first trial (Figure 7a) was very short ($t = 184$ ms) and thus the first target selection map of the SceneWalk StartMap model is biased strongly toward the center. The activations on this map translate into probabilities for being “fixated” by the model. Thus trials with short initial saccade latencies produce many fixations close to the image center. The second trial (Figure 7b) had an initial saccade latency of 1484 ms (1000-ms ms pretrial fixation time + 484 ms after the fixation cross vanished), which is enough to replace the central activation map with the empirical density map of the image multiplied with a Gaussian around the starting position. After a long saccade latency, this map is roughly the same map as the original SceneWalk model without an explicit center bias produces and leads to mean fixation positions further away from the image center.

We simulated saccadic sequences from the SceneWalk StartMap model with the same starting positions, number of fixations, and fixation durations as observed empirically. The temporal evolution of the DTC of Experiment 1 for different pretrial fixation times for the SceneWalk StartMap model is shown in Figure 8a. The SceneWalk StartMap model took the initial saccade latency after image onset into account, which produced a qualitatively similar pattern for the different pretrial fixation times as seen in the data. The qualitative progression for most pretrial fixation times was similar to what was observed empirically. It is eminent though that the central fixation tendency produced by the model was too weak when compared with the data. This was probably a result of the method and the fixations used for the parameter estimation (see Appendix).

We also evaluated the relation between latencies of the first saccade and DTC of the second fixation (Figure 8b). This influence was also visible in the SceneWalk StartMap model, because longer initial saccade latencies led to a less pronounced central activation map. The SceneWalk StartMap model produced a result pattern similar to the empirical data with a similar progression of lines and a differentiation. 
Figure 7. Simulated fixations 1–4 (left to right) of two trials on the same image. (a) A pretrial fixation time of 0 ms and a saccade latency of 184 ms create an attention map for the first saccade target biased toward the center. This leads to fixations close to the center. (b) A pretrial fixation time of 1000 ms and saccade latency of 1484 ms create an attention map for the first saccade target without a central bias. The initial attention map in this trial roughly represents the empirical density map of the image multiplied with a Gaussian around the starting position.

Figure 8. (a) Distance to image center over time for the empirical data and the SceneWalk StartMap model for the different pretrial fixation times in Experiment 1. (b) Influence of the initial saccade latency on the distance to image center on the second fixation for the empirical data and the SceneWalk StartMap model in Experiment 1.
between pretrial fixation times. However, the early CFB on the second fixation was too small in all experiments, i.e., the distance to center in all simulations was too large.

**Discussion**

Adjusting an existing model of saccade generation with an initial central activation map whose influence declines with increasing saccade latency can reasonably explain the central fixation bias. The SceneWalk StartMap model qualitatively replicated differences in DTC curves between pretrial fixation times, and replicated saccade latency effects on DTC of the second fixation. However, the CFB from our simulations was too weak, which is probably a result from the methods used for estimating the parameters (see Appendix). Replicating our empirical findings was accomplished by assuming that the central fixation bias is a result of a default activation in the center of a suddenly appearing stimulus, which is gradually replaced by a content-driven activation.

**General discussion**

During scene viewing the eyes have a strong tendency to fixate near the center of an image, which potentially masks other bottom-up and top-down effects of saccadic target selection. In a previous study (Rothkegel et al., 2016) with starting positions near the image border and an experimentally delayed first saccade after the onset of an image we observed a considerable reduction of the central fixation bias (CFB; Tatler, 2007). Here, we investigated this reduction in four scene-viewing experiments. We manipulated starting positions and the latency of the initial saccade. Different to the original scene-viewing paradigm, where participants start exploration immediately after image onset, we delayed the initial saccadic response by instructing participants to start exploration only after disappearance of a fixation marker. As a measure of the central fixation bias we computed the distance to center (DTC) of the eyes over time. In all experiments the disappearance of a fixation marker 125 ms after image onset led to an early reduction of the CFB in comparison to trials where the fixation marker disappeared simultaneously to the image onset (original scene-viewing paradigm). The earliest pretrial fixation time to produce an influence was measured at 75 ms (see Experiment 3). The reduction of the CFB was particularly pronounced in experiments with pretrial fixation time as a between-subject factor (Experiment 1 and 2). A reduction of the CFB was even visible when participants started observation at the center of an image (Experiment 4). The distance to center of the second fixation was well predicted by the latency of the initial saccade (time from image onset) across experiments. Short saccade latencies led to a strong bias toward the center whereas longer saccade latencies were less systematically directed toward the center. Hence, the latency of the initial response seemed to primarily account for the observed differences of the CFB.

Previous studies have shown that it takes 90 ms on average for the visual input to reach the cortical areas (Clark, Fan, & Hillyard, 1994; Di Russo, Martinez, Sereno, Pitzalis, & Hillyard, 2002) and at least another 60 ms to execute an already programmed saccade (e.g., Findlay & Harris, 1984; Ludwig, Mildinhall, & Gilchrist, 2007). Thus, to plan a saccade to an image-dependent location the latency has to be at least 150 ms. All pretrial fixation times smaller than 125 ms contained trials with initial saccades latencies below 150 ms. Thus implementing a pretrial fixation time of 125 ms and more removed all saccades, which could not have been the result of an image-specific target selection. Additionally, our results and model simulations have shown that the central fixation bias gradually decreases for latencies from 150 ms to 400 ms. Thus we propose a delay somewhere between 125 ms and 250 ms for a strong and reliable reduction of the CFB.

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

We propose another explanation for the early central fixation bias. Our results have shown that the sudden image onset is a dominant contributor to the persistent early central fixation bias. Previous research has shown
that the sudden appearance of a new stimulus captures attention and attracts eye movements, even if it is completely task irrelevant (Theeuwes, Kramer, Hahn, & Irwin, 1998). If, however, this suddenly appearing stimulus appears during a fixation when no saccade is being programmed it does not guide gaze irrespective of the task (Tse, Sheinberg, & Logothetis, 2002). These results can be transferred to our experiment in the following way: The sudden luminance change on the monitor when an image is displayed can be treated as a large object (the image) suddenly appearing. When looking at an object, the eyes usually try to land in the center (Nuthmann & Henderson, 2010) and suddenly appearing objects are being fixated close to the center (Kowler & Blaser, 1995; Richards & Kaufman, 1969).

Thus, if an image suddenly appears and a saccade is planned in parallel, this saccade is a reflexive, stimulus-driven saccade towards the appearing stimulus executed via a subcortical path (Munoz & Everling, 2004; Ottes, Van Gisbergen, & Eggermont, 1985). If some time passes after the sudden onset, before the saccade is executed, a saccade can be planned via a cortical path (Munoz & Everling, 2004; Ottes et al., 1985), targeting a location defined by the image content.

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

Computational models that aim at predicting the allocation of visual attention on an image are based on the extraction of image features (Borji & Itti, 2013; Itti et al., 1998) and top-down cognitive processes (Cerf, Harel, Einhäuser, & Koch, 2008; Navalpakkam, Arbib, & Itti, 2005). These models are evaluated by comparing human fixations with a weighted distribution of different influences (Bylinskii et al., 2015; Borji, Cheng, Jiang, & Li, 2015; Borji & Itti, 2013; Le Meur & Baccino, 2013). Although bottom-up and top-down influences as well as a combination of the two can predict human fixations (Bylinskii et al., 2015), the CFB is a strong predictor that improves goodness-of-fit more than any other single feature (Bylinskii et al., 2015; Judd, Ehinger, Durand, & Torralba, 2009). Thus, saliency models are usually compared with the CFB as a baseline (e.g., Bruce, Wloka, Frosst, Rahman, & Tsotsos, 2015; Clarke & Tatler, 2014; Wilming, Betz, Kietzmann, & König, 2011) and rely heavily on the implementation of a CFB for a good performance (Kümmerer, Wallis, & Bethge, 2015). Because the early CFB during scene viewing seems to be an automated, stereotyped response of the saccadic system to a sudden image onset, it masks bottom-up and top-down factors of saccade target selection and its strength critically depends on the duration of a trial since it primarily affects early fixations. Therefore, a reduction of the CFB during scene viewing, as generated by our paradigm, provides a better understanding of target selection and a more rigorous test of visual attention models than the original scene-viewing paradigm. At the minimum, the latency of the first saccade needs to be taken into account, because it strongly influences subsequent viewing behavior.

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

**Conclusion**

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

**Keywords:** eye movements, dynamic models, visual scanpath, visual attention
Acknowledgments

This work was supported by Deutsche Forschungsgemeinschaft (grants EN 471/13-1 and WI 2103/4-1 to R. E. and F. A. W., resp.). We thank Benjamin Tatler for providing us with the images of natural scenes from his seminal paper about the central fixation bias. We would like to thank Miguel Eckstein for editing our manuscript and Sebastian Pannasch and one anonymous reviewer for their concise reviews, which, in our opinion, increased the quality of the manuscript substantially.

Commercial relationships: none.
Corresponding author: Lars Rothkegel.
Email: lars.rothkegel@uni-potsdam.de.
Address: Department of Psychology, University of Potsdam, Potsdam, Germany.

References


Appendix

SceneWalk model

For our model simulations we took the existing SceneWalk model of saccade generation (Engbert et al., 2015) and extended it to model the early central fixation bias. The SceneWalk model proposes that eye movements are driven by two different time-dependent neural activation maps. An attention map reflects the attentional allocation on the given scene for a specific fixation position. To compute the attention map, first, an intermediate map is computed by multiplying a two-dimensional Gaussian distribution centered at the current fixation position with the empirical saliency map of the image to reflect the reduced processing in the periphery. The influence of attention maps from previous fixations declines over time and thus the previous attention map is increasingly replaced by the map of the new fixation. A second map, the fixation map, memorizes previous fixations and tags visited fixations locations, making them less probable to be fixated again shortly afterward. Thus, this map serves as an inhibition of return mechanism (Itti & Koch, 2001; Klein, 2000). The mechanism to control the dynamics of inhibition, i.e., the fixation map, is equivalent to the mechanism used for the attention map. The attention and inhibition maps prior to the first fixation are set to zero. After computation of the two maps for the current fixation position and duration, they are combined by subtracting the fixation map from the attention map to a target map. After the maps are combined, a target is chosen proportional to the relative activations (Luce, 1959) of the target map. Thus, positions where the fixation map is high whereas the attention map is low are rarely fixated and vice versa. For the interested reader the complete architecture of the model can be found in (Engbert et al., 2015) and a newer version in (Schütz et al., 2017).
SceneWalk StartMap model

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

First, different than in the original SceneWalk model with zero activation across the entire attention map at the beginning of a trial, we used an attention map with higher activations near the center of an image than at the periphery (see Figure 7a). This was motivated by the sudden image onset that may lead to an initial prioritization of central locations. This activation was a two-dimensional Gaussian centered at the image center with two different standard deviations for the horizontal and vertical dimension ($\sigma_x$ and $\sigma_y$). This initial attention map was normalized to a sum of 1.

Second, we realized that the decay of the attention map was too fast during the initial fixation. Therefore, we estimated a new parameter $\rho_2$ that specified the rate of decay during the initial fixation. For all other fixations we used the same decay parameter as during the original simulations (Engbert et al., 2015).

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

$$a(t) = \varphi A_{i,j}(t) + e^{-\rho_2}(a(t) - \varphi A_{i,j}(t)), \quad (2)$$

where $a(t)$ is the attention map at time ($t$) and $A_{i,j}\varphi$ is the empirical density map multiplied with a Gaussian around the starting position $i,j$. The new decay parameters $\rho_2$ controls the speed with which the initial central activation map is replaced. Thus with increasing saccade latency (increasing $t$) the initial central activation map [i.e., $a(0)$] is gradually replaced by the empirical saliency map multiplied with a Gaussian around the starting position ($\varphi A_{i,j}$).

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

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