Fixation durations in natural scene viewing are guided by peripheral scene content

Wolfgang Einhäuser
Physics of Cognition Group, Institute of Physics, Chemnitz University of Technology, Chemnitz, Germany

Charlotte Atzert
Physics of Cognition Group, Institute of Physics, Chemnitz University of Technology, Chemnitz, Germany
Cognitive and Integrative Systems Neuroscience, Philipps-University Marburg, Marburg, Germany

Antje Nuthmann
Perception and Cognition Group, Institute of Psychology, University of Kiel, Kiel, Germany

Fixation durations provide insights into processing demands. We investigated factors controlling fixation durations during scene viewing in two experiments. In Experiment 1, we tested the degree to which fixation durations adapt to global scene processing difficulty by manipulating the contrast (from original contrast to isoluminant) and saturation (original vs. grayscale) of the entire scene. We observed longer fixation durations for lower levels of contrast, and longer fixation durations for grayscale than for color scenes. Thus fixation durations were globally slowed as visual information became more and more degraded, making scene processing increasingly more difficult. In Experiment 2, we investigated two possible sources for this slow-down. We used "checkerboard" stimuli in which unmodified patches alternated with patches from which luminance information had been removed (isoluminant patches). Fixation durations showed an inverted immediacy effect (longer, rather than shorter, fixation durations on unmodified patches) along with a parafoveal-on-foveal effect (shorter fixation durations, when an unmodified patch was fixated next). This effect was stronger when the currently fixated patch was isoluminant as opposed to unmodified. Our results suggest that peripheral scene information substantially affects fixation durations and are consistent with the notion of competition among the current and potential future fixation locations.

Introduction

Two main questions are of interest when studying gaze control during natural scene viewing: *where* in the image do observers fixate, and *when* do they proceed and shift their gaze to the next location? The "where" question regarding fixation *locations* has been studied extensively both experimentally (Williams & Castelhano, 2019, for review) and by computational modeling (Borji & Itti, 2013; Tatler, Hayhoe, Land, & Ballard, 2011, for reviews). By comparison, less research has been conducted on the "when" question, that is the control of fixations *durations* during scene viewing (Nuthmann, 2017, for review). We aimed to help fill this gap by testing the degree to which fixation durations adapt to global scene processing difficulty (Experiment 1), and the degree to which different regions of the visual field influence fixation duration (Experiment 2). In particular, we distinguish scene processing in foveal vision (∼1° to either side of fixation) as opposed to extrafoveal vision, which comprises both the parafovea (from 1°–5° on either side of fixation) and the periphery (>5°).

It has been known for a long time that fixation durations in visual-cognitive tasks vary with processing difficulty (Rayner, 1978). For sentence reading, the various factors that influence fixation durations have been extensively investigated using experimental (Rayner, 1998; Rayner, 2009, for reviews), corpus-analytical (e.g., Kliegl, Nuthmann, & Engbert, 2006), and computational (E-Z Reader: Reichle, Rayner, & Pollatsek, 2003; SWIFT: Engbert, Nuthmann, Richter, & Kliegl, 2005) approaches. The ways in which properties of a fixated word *n* directly influence fixation durations on word *n* can be described as *immediacy effects*. Generally, fixation durations are longer on difficult (e.g., low-frequency) than on easy (e.g., high-frequency) words (e.g., Inhoff & Rayner, 1986; Kliegl et al., 2006). Although the existence of immediacy effects is undisputed, the existence of *parafoveal-on-foveal* effects is controversially discussed (Drieghe, 2011, for review). Parafoveal-on-foveal effects

Citation: Einhäuser, W., Atzert, C., & Nuthmann, A. (2020). Fixation durations in natural scene viewing are guided by peripheral scene content. *Journal of Vision*, *20*(4):15, 1–15, https://doi.org/10.1167/jov.20.4.15.

https://doi.org/10.1167/jov.20.4.15 Received September 28, 2019; published April 24, 2020 ISSN 1534-7362

describe the degree to which properties of a parafoveal word influence fixation durations on the foveal word during reading.[1]

More recently, researchers have investigated global adjustments of fixation durations during scene perception. Examples for global control are effects of viewing task (Mills, Hollingworth, Van der Stigchel, Hoffman, & Dodd, 2011), and the effects image-wide degradations of low-level features have on fixation duration. Fixation durations were found to increase when removing high-spatial frequency information through low-pass filtering (Mannan, Ruddock, & Wooding, 1995), when removing low-spatial frequency information through high-pass filtering (Cajar, Engbert, & Laubrock, 2016), or when removing phase information (1/*f* noise; Kaspar & König, 2011; Walshe & Nuthmann, 2015). Importantly for the present study, fixation durations were increased when color was removed from scene stimuli in both a scene memorization task (von Wartburg et al., 2005), a free viewing task (Ho-Phuoc, Guyader, Landragin, & Guerin-Dugue, 2012), as well as an object-in-scene search task (Nuthmann & Malcolm, 2016). Moreover, Loftus (1985, experiment 5) reported longer mean fixation durations on scenes that were viewed at lower luminance-contrast levels (see also Henderson, Nuthmann, & Luke, 2013). In Experiment 1, we further investigated the role color and contrast play in controlling fixation durations in scene viewing.

The CRISP (Timer **C**ontrolled **R**andom-walk with **I**nhibition for **S**accade **P**lanning) model provided a first theoretical and computational account of fixation durations in scene viewing (Nuthmann, Smith, Engbert, & Henderson, 2010). The model is based on the fundamental principle that moment-by-moment demands on visual and cognitive processing inhibit saccade initiation, and therefore prolong fixation duration. CRISP can account for global adjustments of fixation durations during scene perception, for example, by capturing task-specific influences through differences in parameter settings (scene memorization vs. search: Nuthmann et al., 2010; free-viewing of naturalistic vs. semi-naturalistic videos: Saez de Urabain, Nuthmann, Johnson, & Smith, 2017; scene viewing vs. reading: Nuthmann & Henderson, 2012). In the model, saccade programming is completed in multiple distinct stages of processing. This multistage saccade programming assumption, which CRISP shares with other models of eye-movement control in visual-cognitive tasks (e.g., Engbert et al., 2005; Reichle et al., 2003), is motivated by findings from double-step experiments (Becker & Jürgens, 1979; Walshe & Nuthmann, 2015).

Double-step experiments have demonstrated that it takes a non-negligible amount of time to program a saccade. Moreover, visual information reaches the brain with some unavoidable delay (the eye-brain lag, lasting about 60 ms; Reichle & Reingold, 2013). This poses the problem that there is little time available during each fixation to allow for immediate real-time adjustments of fixation durations by the currently foveated stimulus. One way to deal with this apparent paradox is to assume a somewhat weaker coupling between eye-movement programming and processing of the currently fixated stimulus (Engbert et al., 2005; Nuthmann et al., 2010). An alternative, or complementary approach is to acknowledge the role parafoveal vision plays in enabling direct control of fixation durations (see analysis by Reichle & Reingold, 2013, for sentence reading).

Few scene-viewing studies have been designed to distinguish between foveal and parafoveal influences on fixation durations, using different approaches. First, one can assess the degree to which the different regions of the visual field influence fixation duration by utilizing the gaze-contingent moving window paradigm (McConkie & Rayner, 1975), and the moving mask paradigm (Rayner & Bertera, 1979). The idea is to continuously remove (or strongly degrade) scene information in a selected region of the visual field (e.g., the fovea) to test whether this affects fixation duration. Results from Nuthmann (2013, 2014) suggest that visual information within both foveal and parafoveal vision can influence fixation duration.

A related approach is to selectively manipulate the presence or absence of a specific low-level feature inside and outside a gaze-contingent moving window. With particular relevance to the present work, in one of these studies the availability of color information was manipulated (Nuthmann & Malcolm, 2016). Scene images were presented in full color, with color in the periphery and gray in central vision (i.e., in the fovea and parafovea), gray in the periphery and color in central vision, or in grayscale. Selectively removing color from either central vision or peripheral vision led to increased fixation durations, suggesting that color information in both central and peripheral vision plays a critical role in regulating fixation durations.

In another set of studies, high-pass or low-pass filters were applied to either central or peripheral regions of the visual field during viewing of color (Laubrock, Cajar, & Engbert, 2013) or grayscale (Cajar, Schneeweiß, Engbert, & Laubrock, 2016) scenes; in additional experiments, filter levels and sizes were manipulated (Cajar, Engbert, et al., 2016). The main hypothesis was that scene processing should be most difficult with central low-pass and peripheral high-pass filtering, as these conditions strongly attenuate the critical spatial frequencies for foveal analysis (high spatial frequencies) and peripheral target selection (low spatial frequencies), respectively. Therefore fixation durations were expected to be prolonged in these conditions. However, the inverse pattern was consistently found: mean fixation durations increased most with central high-pass and peripheral low-pass filtering. At a more general level, this means that

observers do not always extend their fixation durations in conditions of increased processing difficulty (Cajar, Engbert, et al., 2016).

Finally, analyzing a large corpus of eye movements during three different scene-viewing tasks, Nuthmann (2017) investigated immediacy effects and parafoveal-on-foveal and/or successor effects[2] of local image statistics on fixation durations. To test the local influence of visual image features, circular image patches with a radius of 1°, approximating foveal vision, were centered on each fixation point. Importantly, in the memorization and preference tasks (but not for the visual search task), some evidence for successor effects emerged, such that some image characteristics of the upcoming location $n + 1$ influenced how long the eyes stayed at the current location n (see also Tatler, Brockmole, & Carpenter, 2017).

Existing models of eye-movement control have treated temporal ("when?") and spatial ("where?") aspects of gaze control independently (e.g., Findlay & Walker, 1999; Nuthmann et al., 2010). However, already decades ago it has been observed that fixation probability and fixation duration are closely related to each other (Buswell, 1935). More recently, it has been shown that fixation probability predicts fixation duration during scene viewing (Einhäuser & Nuthmann, 2016). Challenging the assumption of separate mechanisms for selection in space and time, Tatler et al. (2017) introduced the LATEST (Linear Approach to Threshold Explaining Space and Time) model, which utilizes a single decision mechanism to explain both when and where we look in scenes. Interestingly, in this model fixation durations predict fixation locations (rather than the other way around, as proposed by Einhäuser and Nuthmann, 2016). Moreover, LATEST incorporates information processing both at the fovea (i.e., at the current fixation) and in peripheral vision (i.e., at the location of the next fixation) to predict saccadic decision times.

In the present study, we used a two-step experimental approach to test the relative influence of visual information at the current and the next fixation location on fixation durations during scene viewing. In the first step, we re-examine the hypothesis that a global reduction of the available visual information yields prolonged fixation durations. In Experiment 1, we tested this hypothesis by presenting pictures of natural scenes in their unmodified form along with versions that were reduced in contrast and/or saturation. We expected reduced contrast and/or color to prolong fixation durations during scene inspection.

In the second step, we investigated two possible sources for this slow-down. First, reducing the available visual information and thereby increasing processing difficulty at the current fixation location may prolong fixation durations. Second, reduced informativeness of potential future fixation locations may delay the decision to leave the current fixation location, thereby also increasing fixation duration. In Experiment 2, we distinguished between these alternatives by using hybrid stimuli in which we manipulated the available visual information differently in alternating patches. Specifically, unmodified patches alternate with patches from which luminance information has been removed (isoluminant patches). This results in a checkerboard-like pattern of alternating isoluminant and unmodified patches (Figure 1). We note that setting the luminance-contrast to zero also eliminates luminance edges, and therefore reduces higher-level feature content. For isoluminant patches, the extraction of information will be more difficult, which is why we conceptualize this manipulation as an increase in processing difficulty. By moving their eyes over the scene, observers will generate observations for the four (2×2) possible combinations of scene degradation at the current fixation (unmodified vs. isoluminant) and the next fixation (unmodified vs. isoluminant). Our hypotheses on the control of fixation durations pertain to the duration of the current fixation. First, we should observe longer fixation durations on isoluminant than on unmodified patches (immediacy effect). Second, fixation durations should be shorter for unmodified as opposed to isoluminant upcoming patches (parafoveal-on-foveal effect). Third, inspired by research on eye-movement control in reading, we tested a prediction derived from the foveal load hypothesis (Henderson & Ferreira, 1990; Schad & Engbert, 2012) according to which difficulties in foveal processing cause processing load, thereby reducing capacities available for parafoveal (pre)processing. Accordingly, the degradation parafoveal-on-foveal effect from the patch selected for the next fixation should depend on the degradation of the currently fixated patch.

Materials and methods

Participants

Twenty observers (13 women, 7 men; age range: 19–36 years, mean ± *SD*: 24.0 ± 3.4) participated in Experiment 1, with 24 (14 women, 10 men; age range: 19–31 years, mean ± *SD*: 23.5 ± 3.2) in Experiment 2. All had normal or corrected-to-normal vision and normal color vision as assessed by Ishihara plates. Experiments conformed to the Declaration of Helsinki and written informed consent was obtained from all participants. The responsible body (Ethikkommission HSW, TU Chemnitz) ruled that no in-depth ethics evaluation was necessary for this study (case-no: V-192-WET-Szenen-10042017). The number of participants was determined at the time of application with the ethics board; that is, prior to conducting

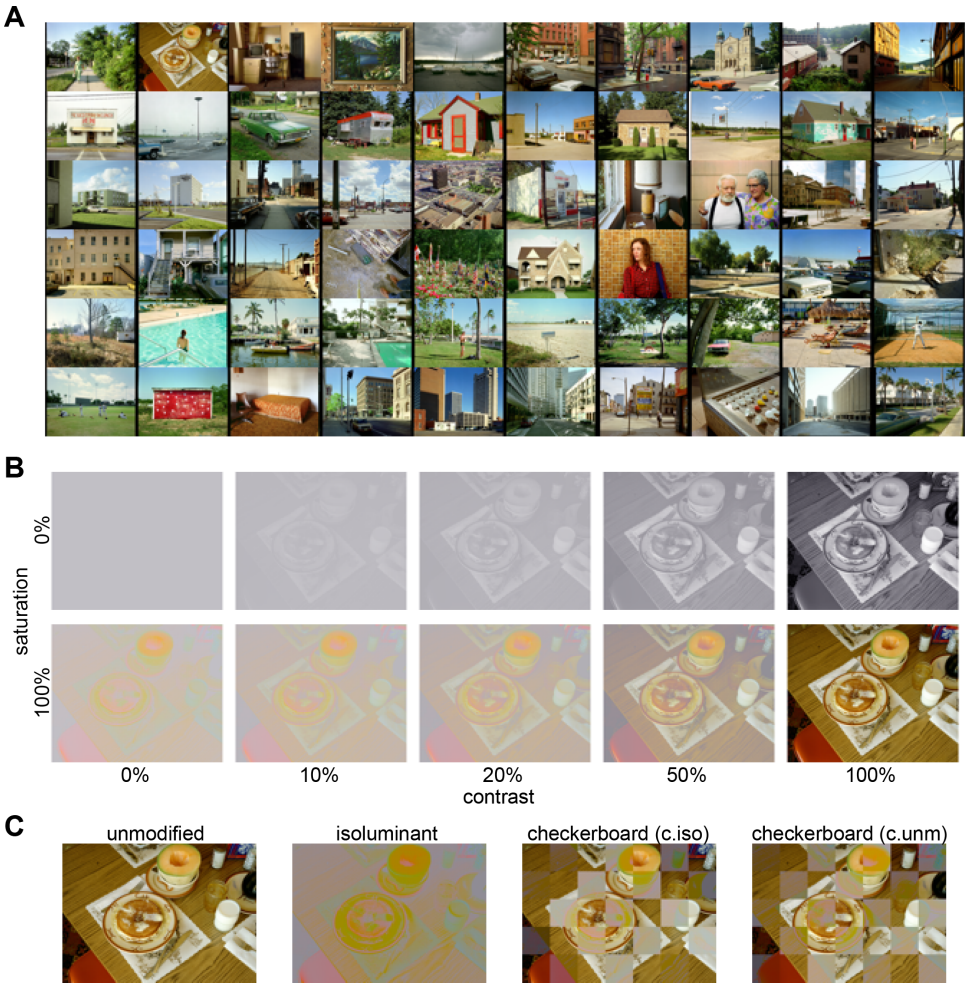

Figure 1. Scene stimuli used in the present study. (A) Thumbnails of the 60 images selected from Shore et al. (2004), reprinted with the permission of the artist. (B) Conditions of Experiment 1, top row: saturation 0, bottom row: full saturation; from left to right: no luminance contrast, 10%, 20%, 50%, and 100% of full image contrast. (C) Conditions of Experiment 2: unmodified image, isoluminant (zero luminance-contrast) image, checkerboard image with center isoluminant, checkerboard image with center unmodified.

the experiment. From data of previous studies (e.g., Stoll et al., 2015), we estimated a number of $N = 20$ participants per experiment for sufficient power; $N = 24$ in Experiment 2 was chosen as the smallest number greater than or equal to 20 that allowed appropriate counterbalancing of stimuli.

Stimuli

Stimuli were generated from 60 images of the Stephen Shore "Uncommon Places" collection (Shore, Tillman & Schmidt-Wulffen, 2004) kindly provided by the artist in digital form at a resolution of 1024 × 768 pixels (Figure 1A). Images were a subset of those used and characterized in earlier eye-tracking studies (Einhäuser, Spain & Perona, 2008). For Experiment 1, images were cropped symmetrically to 952 × 768 pixels to remove black boundaries that are visible in some of the images. For Experiment 2, cropping was to 945 × 675 pixels to allow for segmenting the image into 7 × 5 squares without interpolation. Horizontal cropping was nearly symmetrical (39 pixels at the left, 40 at the right); vertically, double the margin was cropped at the bottom (62 pixels) than at the top (31 pixels).

In Experiment 1, 10 versions of each image were created, yielding 600 stimuli in total. There were two different color conditions: for half of the stimuli, saturation was reduced to 0 (condition "grayscale"), for the other half, saturation was left unchanged (condition "color"). For each of the color conditions, five contrast conditions were created. From the unmodified image's luminance values (as it would appear on the display used) the mean displayable luminance (49.8 cd/m^2) was subtracted, the resulting values multiplied by a factor 0, 0.1, 0.2, 0.5, or 1, and the mean displayable luminance was re-added. This results in stimuli that are reduced in contrast to

0% (isoluminant at 49.8 cd/m^2), 10%, 20%, 50%, or 100% of the original image (Figure 1B). Note that the condition "color/100% contrast" corresponds to the original image, whereas the condition "grayscale/0% contrast" corresponds to an empty screen. As the image database provides no information on the recording conditions, the unmodified images were used as saved by the artist. Hence all modifications were computed relative to the *displayed* unmodified image, using a careful characterization of the screen's chromatic and luminance properties.

In Experiment 2, four versions of each image were used (Figure 1C), leading to 240 distinct stimuli: (a) the image in its original form ("unmodified"), (b) all pixels' luminance set to the mean image luminance ("isoluminant"), and (c) stimuli consisting of isoluminant patches alternating with unmodified patches. The isoluminant patches in these stimuli were set to the original patch's mean luminance. Patches were 135 × 135 pixels wide. As the alternation is analogous to the black/white alternation of a checkerboard, we will refer to these stimuli as "checkerboard" stimuli and to the patches as "checks." As this modification tiles the checkerboard images in 7 × 5 checks, there are two different versions: checkerboard stimuli with an isoluminant central patch ("checkerboard center isoluminant [c. iso.]") are distinguished from those with an unmodified central patch ("checkerboard center unmodified [c. unm.]"). Note that the "color/0% contrast" condition of Experiment 1 deviates slightly from the isoluminant condition in Experiment 2, as the former is set to the mean displayable luminance, whereas the latter is set to the mean image luminance.

Isoluminance was chosen as manipulation because (a) the removal of luminance information had been known to influence fixation behavior (e.g., scan paths, Harding & Bloj, 2010); (b) luminance affects fixation durations in natural scenes above and beyond other features (Nuthmann, 2017); and (c) unlike spatial filtering, the modification acts locally, that is, at the pixel level, such that no cross-talk across spatial scales is to be expected. With Experiment 1, we verified the effectiveness of the modification for whole scenes, and with Experiment 2 we distinguished between effects of current and future fixation locations.

Setup

Stimuli were presented centrally on a VIEWPixx/3D full monitor (VPixx Technologies Inc., Saint-Bruno, QC, Canada) running at a frame rate of 120 Hz and a resolution of 1920 × 1080 pixels. The nonimage background was set to 49.8 cd/m^2 ("gray"), maximum luminance ("white") was 99.6 cd/m^2. Observers were seated at 57-cm distance from the screen, where their head was stabilized with a padded forehead rest and chin rest. Stimuli spanned approximately 25.6° × 21.2° (width × height) in Experiment 1, and 25.4° × 18.7° in Experiment 2. Each "check" of the checkerboard stimuli in Experiment 2 spanned approximately 3.7° in each dimension.

Eye movements were recorded monocularly at 1000 Hz with an Eyelink-1000 Plus (SR Research, Ottawa, ON, Canada) infrared eye-tracking device. In all but one observer, data were recorded from the left eye; for one observer in Experiment 2 the right eye was used instead. Blinks and saccades were detected by using the eye tracker's built-in functions with saccade thresholds of 35°/s for velocity, and 9500°/s^2 for acceleration. A 13-point calibration and validation procedure was applied at the beginning of each block and whenever an initial fixation (see Procedure later) failed for technical reasons.

For stimulus preparation, stimulus presentation, eye-tracker control, and data analyses MATLAB (The MathWorks, Natick, MA) was used, including the Psychophysics Toolbox (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007), which incorporates the Eyelink Toolbox extensions (Cornelissen, Peters & Palmer 2002).

Procedure

In Experiment 1, observers started each trial with a central fixation on a black fixation cross on a gray background. If an observer failed to fixate within 1° of visual angle (36 pixels) of the cross for at least 300 ms within 3.5 seconds, the eye tracker was recalibrated, and the trial restarted. Otherwise, image onset was after 300 ms of fixation. The image was presented for 4 seconds, followed by a gray screen. After approximately 100 to 250 ms the next trial started. Each image was used in each condition exactly once (600 trials in total). The 600 trials were split in 10 blocks of 60 stimuli. In each block, each image was used exactly once, and each condition occurred six times. The presentation order within each block was random.

In scene-viewing studies, participants are oftentimes instructed to simply look at the images (Ho-Phuoc et al., 2012; Saez de Urabain et al., 2017). However, this free-viewing task is conceptually problematic as different observers could interpret the task differently (Tatler et al., 2011). Therefore we instructed our participants to "study the images carefully" (cf. Einhäuser & König, 2003), as this is a fairly general instruction, promoting the employment of broadly similar viewing strategies between participants (Nyström & Holmqvist, 2008). Moreover, participants were informed that they were allowed to move their eyes naturally once the fixation cross disappeared and the image came on.

The procedure of Experiment 2 was identical with the following exceptions. First, there were only 240 different stimuli, hence there were only four blocks of 60 trials each. Second, the presentation duration of each stimulus was extended to 8 seconds. The assignment of images and conditions per block was such that each image occurred only once per block and each condition 15 times per block. Across observers, the assignment of image/condition to block was counterbalanced, such that for each image each order of conditions occurred exactly in one observer. As there are 24 (= 4!) possible orders for four conditions, this requires the number of subjects to be an integer multiple of 24.

Analyses

In both experiments, two dependent variables were analyzed. First, we considered the latency of the first saccade, defined as the time from image onset to the end of the initial central fixation. Second, we analyzed the duration of all subsequent fixations ending prior to image offset (i.e., the last fixation starting during image display and ending thereafter was excluded). Considering that the distributions of latencies and fixation durations can be substantially skewed (e.g., Nuthmann et al., 2010, for fixation durations), the median latency across all images per subject and condition, and the median fixation duration across all images and fixations per subject and condition were used as measures.

For Experiment 1, a 2 × 5 repeated-measures analysis of variance (ANOVA) with factors color (two levels: color, grayscale) and contrast (five levels: 0, 10%, 20%, 50%, 100%) was computed for each dependent variable. When significant interactions were observed, follow-up *t*-tests were conducted at each level of contrast. By making the scene image disappear when combining no color (grayscale) with zero contrast, this condition is qualitatively very different from the others. Moreover, image onset and hence the latency of the first saccade may be regarded ill-defined in this case. Therefore we ran complementary 2 × 4 ANOVAs in which the zero-contrast conditions were excluded.

For the data from Experiment 2, analyses focused on the checkerboard stimuli. Here we considered pairs of fixations and refer to the current fixation as fixation n and the next fixation as $n + 1$. Fixations were split into four conditions, depending on the patches that were associated with fixations n and $n + 1$. When analyzing fixation durations, the dependent variable was the duration of fixation n. Latencies were subjected to a 2 × 2 repeated-measures ANOVA with the factors current patch (levels: isoluminant, unmodified) and next patch (levels: isoluminant, unmodified). For fixation durations, the type of checkerboard stimulus (levels: center unmodified, center isoluminant) was an additional factor, leading to a 2 × 2 × 2 repeated-measures ANOVA (note that for the variable latency, the type of checkerboard stimulus is redundant to the current patch, as the initial fixation by design is always on the central patch). In addition, latencies, as well as fixation durations, were compared between completely unmodified and complete isoluminant stimuli by means of paired *t*-tests.

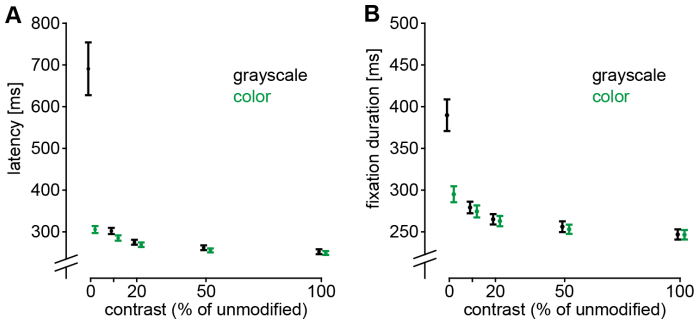

Figure 2. Results of Experiment 1. (A) Median latency of the first saccade after stimulus onset. (B) Median fixation duration (excluding initial central fixation); error bars denote means and *SEM* across observers.

Data availability

All eye-tracking data are available at https://doi.org/10.5281/zenodo.3646545.

Results

Experiment 1

Latency of the first saccade

There was a significant main effect of contrast, $F(4,76) = 49.9$, $p < 0.001$, with longer latencies for lower levels of contrast (Figure 2A). There was also a significant main effect of color: the very first saccade was launched, on average, with a longer latency when color was removed from the scene image, $F(1,76) = 49.6$, $p < 0.001$. The interaction between contrast and color was also significant, $F(4,76) = 42.8$, $p < 0.001$. According to follow-up tests, differences between color stimuli and grayscale stimuli were only observed for low contrast levels [$t(19) = 6.71$, $p < 0.001$ for zero contrast; $t(19) = 3.24$, $p = 0.004$ for 10% contrast], but not for higher contrasts, all $t(19) < 1.89$, all $p > 0.07$. Importantly, when excluding the zero-contrast conditions, the main effects of contrast, $F(3,57) = 69.2$, $p < 0.001$, and color, $F(1,57) = 18.7$, $p < 0.001$, were still significant, whereas the interaction failed to reach statistical significance, $F(3,57) = 2.32$, $p = 0.09$.

Fixation durations

For fixation durations, there was a significant main effect of contrast, $F(4,76) = 62.5$, $p < 0.001$, with longer fixation durations for lower levels of contrast (Figure 2B). There was also a significant main effect of color, $F(1,76) = 33.2$, $p < 0.001$, with longer fixation durations for grayscale as compared with color scenes. The contrast × color interaction was also significant, $F(4,76) = 42.8$, $p < 0.001$. According to follow-up tests, the difference between color conditions was statistically significant for the zero-contrast condition, $t(19) = 5.61$, $p < 0.001$; at 10% contrast the effect of color just failed to be significant, $t(19) = 2.07$, $p = 0.052$, whereas it disappeared for higher contrast levels, all $t(19) < 0.87$, all $p > 0.39$. When excluding the zero-contrast conditions, the significant main effect of contrast was preserved, $F(3,57) = 57.8$; $p < 0.001$. The main effect of color was just significant, $F(1,57) = 4.48$, $p = 0.048$, whereas the contrast × color interaction was not significant anymore, $F(3,57) = 0.74$, $p = 0.53$.

Experiment 2

Probability of fixating an isoluminant patch

In Experiment 2, we introduced "checkerboard" stimuli, in which half of the "checks" contain the original image (unmodified checks), whereas the other half contains a version deprived of luminance information (isoluminant patches). For such checkerboard images, the type of patch that is associated with the initial fixation is determined by the experimental condition (center isoluminant vs. center unmodified). For any subsequent fixations, observers can select either isoluminant or unmodified patches. Isoluminant patches were selected significantly more often when viewing center-isoluminant stimuli (38.9%, $SD = 4.4\%$) as compared with center-unmodified stimuli (37.3%, $SD = 5.5\%$); $t(23) = 2.94$, $p = 0.007$. This small but reliable difference calls for a separate analysis of the two stimulus types. At the same time, these unconditional probabilities show that observers prioritized unmodified patches over isoluminant patches.

Probability of transitions

For the subsequent analysis of initial saccade latencies and fixation durations it is important to know whether the probability of selecting an isoluminant or unmodified patch with fixation $n + 1$ depends on whether fixation n is on an isoluminant or unmodified patch. When the current fixation n was on an unmodified patch, the probability that the next fixation $n + 1$ fell on an isoluminant patch was 35.3% (5.4%) and 34.5% (5.5%) for center-isoluminant and center-unmodified stimuli, respectively (Figure 3). Consequently, the probability that the next fixation $n + 1$ fell on an unmodified patch, if the current fixation n fell on an unmodified patch was 64.7% (100%-35.3%) for center-isoluminant stimuli and 65.5% (100%–34.5%) for center-unmodified stimuli. We compared these conditional probabilities p(n+1|n), that is, the probability to fixate an isoluminant/unmodified patch at fixation $n + 1$ given the type of patch fixated at fixation n, to the unconditional probabilities given earlier. The conditional probability p(isoluminant|unmodified) for the two stimulus types, that is, 35.3% and 34.5%, was lower than the unconditional probability p(isoluminant), that is, 38.9% and 37.3%, $t(23) = 8.01$, $p < 0.001$ and $t(23) = 6.41$, $p < 0.001$.

When fixation n was on an isoluminant patch, fixation $n + 1$ landed on an isoluminant patch with probability 44.9% (4.2%) for center-isoluminant stimuli, and 42.4% (6.0%) for center-unmodified stimuli (Figure 3). These conditional probabilities p(isoluminant|isoluminant) were higher than the unconditional probabilities, $t(23) = 8.73$, $p < 0.001$ and $t(23) = 6.47$, $p < 0.001$. Taken together, if an unmodified patch was fixated at fixation n, the probability to fixate an isoluminant patch at fixation $n + 1$ was higher than the unconditional probability of fixating an isoluminant patch in general. Conversely, if an unmodified patch was fixated at fixation n, the probability to fixate an isoluminant patch at fixation $n + 1$ was lower than the unconditional probability of fixating an isoluminant patch. Put succinctly, observers were more likely to stay on a patch of the same type (either unmodified or isoluminant) than expected by the unconditional probabilities. This includes transitions within the same patch, which accounted for 28.1% ($SD = 6.2\%$) of transitions, with no difference between stimulus types, $F(3,69) = 0.80$, $p = 0.50$.

There was a sufficient number of transitions of each type to allow for splitting analyses by transition. This is more evident if the transitions are expressed in fractions of overall fixations. Excluding the initial fixation, this yields: unmodified to unmodified: 39.8% and 41.3% for center-isoluminant and center-unmodified stimuli, respectively; isoluminant to unmodified: 21.3%, 21.3%; unmodified to isoluminant: 21.4%, 21.4%, and isoluminant to isoluminant: 17.6%, 16.1%.

Saccade length and direction

The checkerboard-like alternation of unmodified and isoluminant patches introduces artificial boundaries, which in itself may add to scene structure. We tested whether these artificial boundaries influenced basic oculomotor measures of spatial selection, that is, saccade length and saccade direction. Averaged across observers and conditions, we observed a median saccade length of 3.1° ($SD = 0.9°$). Importantly,

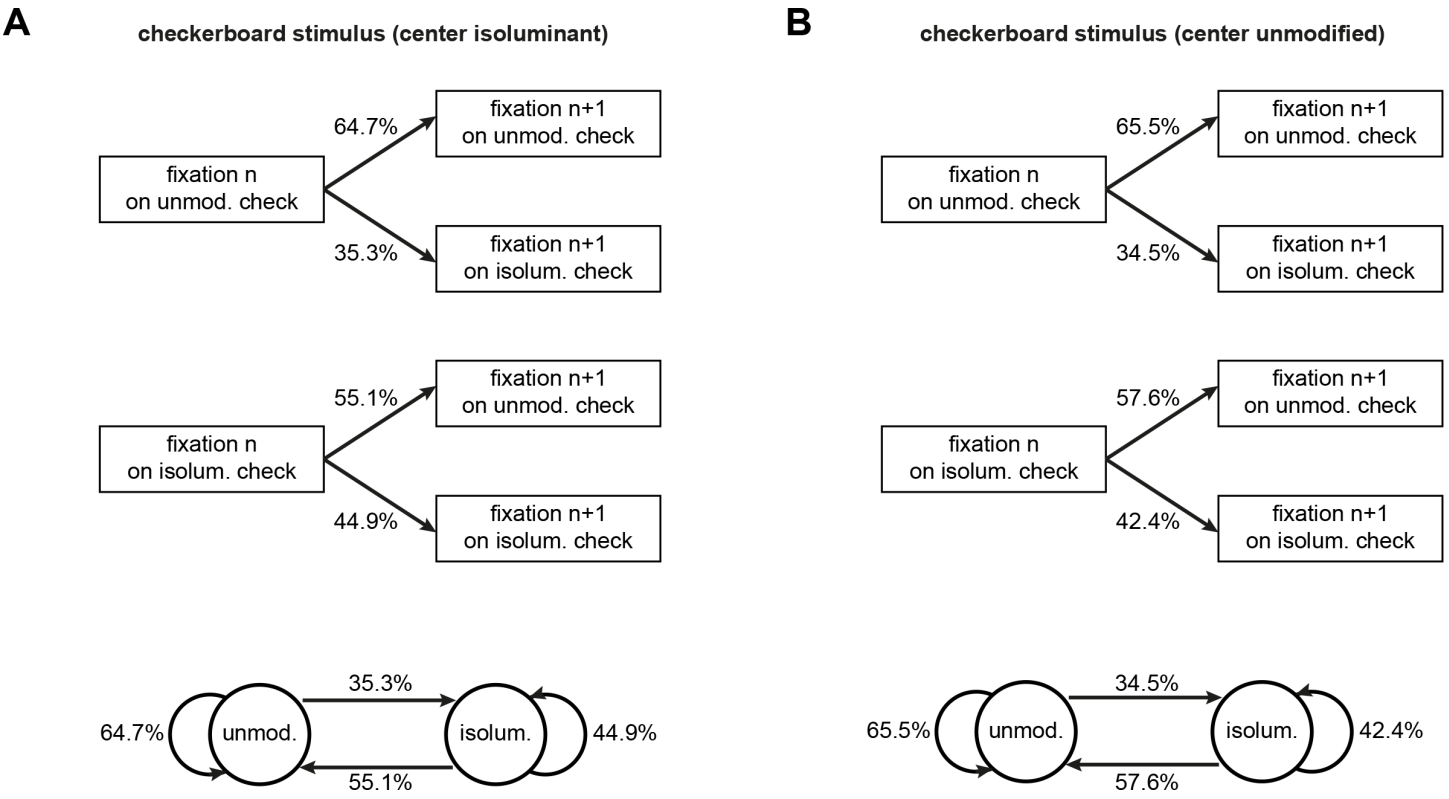

Figure 3. Transition probabilities. For a given type of patch (unmodified/isoluminant), numbers on arrows indicate the probability to transit to either the same type or the other type (i.e., the conditional probabilities p(n+1|n) are provided, thus outgoing arrows add up to 100% at each state). (A) center-isoluminant stimuli, (B) center-unmodified stimuli.

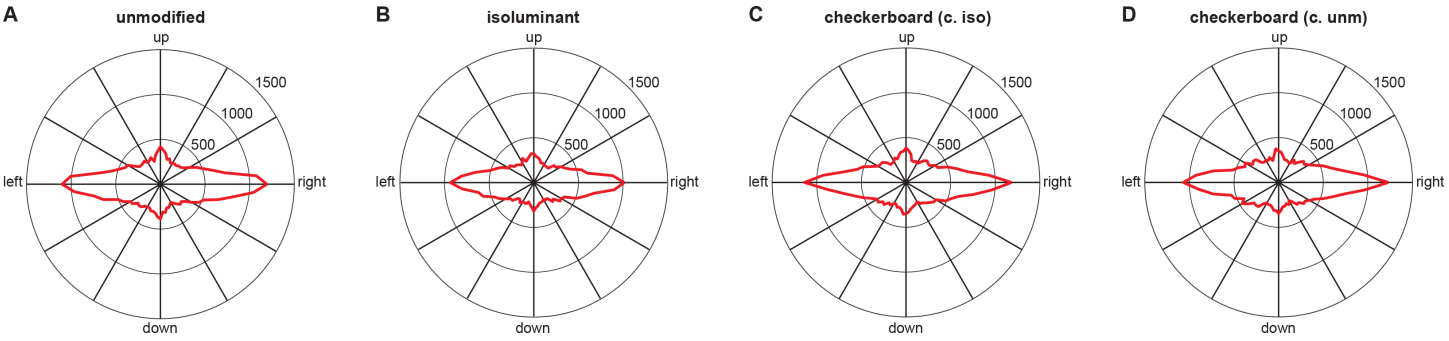

Figure 4. Distribution of saccade directions. Histograms of saccade directions aggregated across observers for the four different stimulus conditions: (A) unmodified stimuli, (B) isoluminant stimuli, (C) center-isoluminant checkerboard stimuli, (D) center-unmodified checkerboard stimuli. The radii represent the absolute number of observations. c. iso, checkerboard center isoluminant; c. unm. checkerboard, center unmodified.

median saccade length did not vary as a function of stimulus type, $F(3,69) = 0.26$, $p = 0.85$. Moreover, we analyzed the angular direction of saccades separately for each condition. All possible directions were divided into 72 bins of 5° each, and for each bin the number of saccades was summed across all observers. The resulting histograms (Figure 4) show the expected horizontal bias, with no substantial differences between conditions. In sum, we see little influence of the checkerboard patterns on spatial properties of saccades.

Latency of the first saccade

Consistent with the results of Experiment 1, the latency of the first saccade (Figure 5A) was significantly shorter for completely unmodified stimuli (mean ±

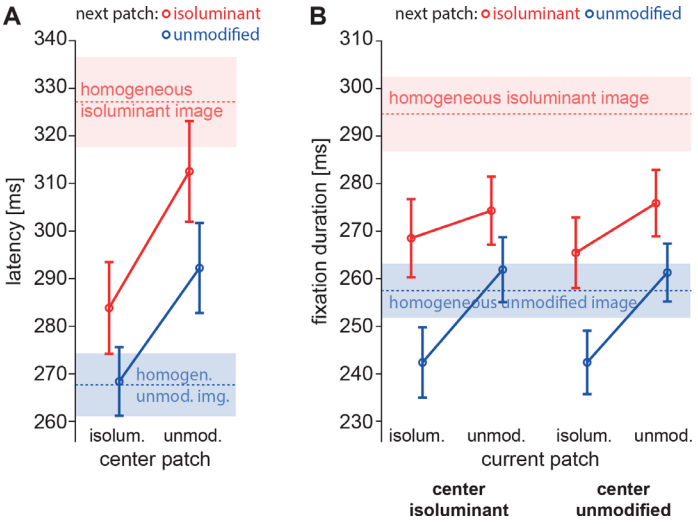

Figure 5. Results of Experiment 2. (A) Median latency of first saccade after stimulus onset split by image type (starting on isoluminant or unmodified central patch) and target (ending on patch of either type, as coded by line color). Horizontal lines denote latency for completely unmodified and completely isoluminant stimuli. (B) Median fixation duration (excluding the initial central fixation) split by image type. The *x*-axis denotes the type of patch the current fixation (*n*) is on, the line color the type of subsequent patch (fixation $n + 1$). Horizontal lines denote fixation duration for completely unmodified and completely isoluminant stimuli. In both panels, error bars and shaded areas denote ±1 *SEM* across observers.

SD: 268 ± 33 ms) than for completely isoluminant stimuli (327 ± 46 ms), $t(23) = -12.6$, $p < 0.001$. For the checkerboard stimuli, there were significant main effects for the type of the initial central patch, $F(1,23) = 61.7$, $p < 0.001$, and for the type of the subsequent patch, $F(1,23 = 18.8$, $p < 0.001$), but no interaction between the two, $F(1,23) = 0.30$, $p = 0.59$. Interestingly, an unmodified central patch was associated with *longer* latencies than an isoluminant central patch. Conversely, latencies were shorter when the first saccade was sent to an unmodified patch as opposed to an isoluminant one.

Fixation durations

Median fixation durations (Figure 5B) were significantly shorter for completely unmodified stimuli (258 ± 28 ms) than for completely isoluminant stimuli (295 ± 38 ms), $t(23) = -8.43$, $p < 0.001$, which is consistent with Experiment 1. For the checkerboard stimuli, there was a significant main effect of the patch on which the current fixation fell, $F(1,23) = 30.4$, $p < 0.001$, with longer fixation durations for unmodified than for isoluminant patches. There was also a significant main effect for the type of patch selected for the next fixation on the duration of the current fixation, $F(1,23) = 81.4$, $p < 0.001$, with the eyes fixating longer on the current patch when the upcoming patch was isoluminant instead of unmodified. The interaction between current and next patches was also significant, $F(1,23) = 6.08$, $p = 0.02$. As there was neither a main effect of stimulus type (unmodified vs. isoluminant center), $F(1,23) = 0.14$, $p = 0.72$, nor any two- or three-way interactions involving the factor stimulus type, all $F(1,23) < 0.67$, all $p > 0.42$, data for all follow-up analyses were aggregated over the two stimulus types.

The interaction between the factors current patch and next patch allowed for follow-up analyses of the pairwise differences between the four (2 × 2) combinations of current and next patch. As mentioned earlier, we observed an inverted immediacy effect (longer, rather than shorter, fixation durations on unmodified patches) along with a parafoveal-on-foveal effect (shorter fixation durations for unmodified upcoming patches), which was stronger when the currently fixated patch was isoluminant as opposed to unmodified. Shortest durations (243 ± 34 ms) were observed when a fixation on an isoluminant patch was followed by a fixation on an unmodified patch. The median fixation duration in this "isoluminant to unmodified" (n to $n + 1$) condition was significantly shorter than fixation durations on completely unmodified images, $t(23) = -5.07$, $p < 0.001$. Compared with the "isoluminant to unmodified" condition, fixations on unmodified patches followed by fixations on unmodified patches were significantly prolonged (261 ± 31 ms); $t(23) = 4.84$, $p < 0.001$. The median fixation duration in this "unmodified to unmodified" condition was statistically indistinguishable from fixations on completely unmodified images, $t(23) = 1.40$, $p = 0.17$. Compared with the "unmodified to unmodified" condition, fixations on isoluminant patches followed by isoluminant patches (isoluminant to isoluminant) had numerically, but not statistically, longer durations (268 ± 38 ms); $t(23) = 2.03$, $p = 0.054$. Longest fixation durations were observed for fixations on unmodified patches followed by an isoluminant patch (275 ± 35 ms). The median fixation duration in this "unmodified to isoluminant" condition was significantly longer than in the "isoluminant to isoluminant" condition, $t(23) = 2.26$, $p = 0.03$, but it was significantly shorter than fixation durations on completely isoluminant stimuli, $t(23) = -4.80$, $p < 0.001$.

Discussion

In two experiments we investigated factors that influence how long the eyes remain fixated on a particular location when viewing images of real-world scenes. In Experiment 1, we tested the degree to which

fixation durations adapt to global scene processing difficulty by manipulating the saturation and contrast of the entire scene. In Experiment 2, we used checkerboard stimuli to manipulate foveal and extrafoveal processing load. As a key finding, fixation durations were affected more strongly by the information content of the upcoming (extrafoveal) patch than the current (foveal) patch.

The majority of research on eye movements during scene perception has focused on the "where" decision regarding the target location for the next saccade (Borji & Itti, 2013; Tatler et al., 2011; Williams & Castelhano, 2019, for reviews). More recently, there has been a growing interest in the "when" decision regarding the control of fixation durations (Nuthmann, 2017, for review). Fixation durations provide a good moment-to-moment indicator of visual-cognitive processing (Rayner, 1998, for review). Supporting this assumption, fixation durations during scene viewing have been shown to globally adjust to overall processing difficulty. Specifically, image-wide degradations of low-level features have been shown to prolong fixations. For example, fixation durations were prolonged when color was removed from scene stimuli (Ho-Phuoc et al., 2012; Nuthmann & Malcolm, 2016; von Wartburg et al., 2005). Moreover, longer mean fixation durations were observed when scenes were viewed at lower luminance-contrast levels (Henderson et al., 2013; Loftus, 1985, experiment 5).

In Experiment 1, we extended this research by independently manipulating the saturation and contrast of naturalistic scenes. For the very first saccade on the image, we observed longer latencies for grayscale than for color scenes (cf. Nuthmann & Malcolm, 2016) and longer latencies for lower levels of contrast. For subsequent fixations, we observed longer fixation durations for lower levels of contrast, and longer fixation durations for grayscale than for color scenes, with subtle differences for specific combinations of saturation and color. The pattern of results is largely consistent with the aforementioned literature. Collectively, the results of Experiment 1 showed that fixation durations were globally slowed as visual information became more and more degraded, making scene processing increasingly more difficult.

In Experiment 2, we followed up on these results by distinguishing between foveal and extrafoveal influences on fixation duration. To this end, we created checkerboard stimuli by superimposing the scenes used in Experiment 1 by a grid (Figure 1C). Local image feature values are bound to show variability across grid cells, whereas some of these features are systematically related to fixation probability (Nuthmann & Einhäuser, 2015). Given this natural variability, we chose a strong experimental manipulation by alternating unmodified patches with isoluminant patches for which the luminance contrast was reduced to zero.

By design, the two checkerboard stimulus types were complementary to each other - where one had an isoluminant patch the other had an unmodified one. Hence each point of any given image contributed once to an unmodified, once to an isoluminant patch, such that it is unlikely that a specific image feature biases the results. Moreover, the size of grid cells (approximately 3.7° × 3.7°) was chosen such that—relative to the currently fixated patch—any neighboring patches or cells would be situated outside foveal vision, in most cases.

Analyses of fixation probability and transition probabilities (fixation $n \rightarrow$ fixation $n + 1$) showed that observers prioritized unmodified patches over isoluminant patches, suggesting that eye guidance was biased toward more informative scene regions. Moreover, observers were more likely to stay on a patch of the same type than expected by the unconditional probabilities. For analyses of initial saccade latencies and fixation durations, the data were split into four conditions, contingent on the type of patch that was associated with the current fixation n and the next fixation $n + 1$.

Depending on the type of checkerboard stimulus, the very first saccade was launched from a central patch that was either unmodified or isoluminant. According to the processing difficulty hypothesis, an unmodified central patch should be associated with shorter latencies than an isoluminant central patch. Interestingly, the opposite was found. Moreover, latencies were shorter when the first saccade was sent to an unmodified patch as opposed to an isoluminant one. These data suggest that saccade latencies do not always increase with foveal processing difficulty, and that they are not only influenced by scene processing in foveal vision but also by extrafoveal processing.

For the checkerboard stimuli, the fixation-duration pattern was qualitatively similar to the latency pattern. We observed a paradoxically inverted immediacy effect (longer, rather than shorter, fixation durations on unmodified patches), an orthodox parafoveal-on-foveal effect in the expected direction (i.e., shorter fixation durations for unmodified upcoming patches), as well as an interaction. The type of checkerboard stimulus, which affected initial saccade latency, ceased to have an effect on subsequent fixation durations.

One reason to conduct Experiment 2 was to test hypotheses on what drives the effects of global image-wide degradations on fixation duration observed in Experiment 1. According to the present data, prolonged fixations at reduced image contrast do not result from the need for more inspection time at lower contrast due to slower visual processing at the currently fixated location. At face value, the pattern of results obtained in Experiment 2 suggests that the difference between isoluminant and unmodified scenes observed in Experiment 1 was determined by scene content at

the next fixation location, whereas the currently fixated location—if anything—counteracted this effect.

The inverted immediacy effect observed in Experiment 2 suggests that observers invested more (less) processing time when the available visual information at fixation could (could not) be efficiently used for gaze control (cf. Cajar, Engbert, et al., 2016). Interestingly, the parafoveal-on-foveal effect was larger in size than the immediacy effect. This means that fixation durations were more strongly affected by differences in processing difficulty in the periphery than at fixation, challenging the widely held assumption that foveal processing plays a dominant role in controlling fixation duration (cf. Cajar, Engbert, et al., 2016). According to the foveal load hypothesis (e.g., Schad & Engbert, 2012, for sentence reading), increased foveal load should not only be associated with longer foveal fixation duration but also with a reduced or absent parafoveal-on-foveal effect. However, the opposite was found: isoluminant patches at fixation were associated with shorter foveal fixation duration and a larger parafoveal-on-foveal effect.

We note that there are three complementary approaches for investigating the interplay between foveal and parafoveal processing. First, researchers may use scene stimuli without any modifications. Although this keeps the scenes naturalistic, results are necessarily correlative in nature (e.g., Nuthmann, 2017). Second, one may change the visual stimulus dynamically depending on the currently fixated location using gaze-contingent manipulations (e.g., Laubrock et al., 2013; Nuthmann & Malcolm, 2016). Third, one may keep the temporal structure of the scene intact, which comes at the cost of disrupting the spatial structure. This is the approach taken here. Given the novelty of this approach, we chose a strong experimental manipulation; in particular, it extends beyond the natural variability of luminance in the scene. Whether the observed effects prevail for more subtle experimental modifications will be an interesting issue for future research. Such modifications may, for example, involve variants that leave some of the luminance variability intact or replace the isoluminant patches by other modifications, such as removal of higher-order information and/or semantic content.

Because the guidance of gaze under natural circumstances is tightly linked to the allocation of spatial attention (e.g., Deubel & Schneider, 1996), it is of interest to note that attention allocation is frequently viewed as a result of competition between items in conjunction with a mechanism that controls priority (Desimone & Duncan, 1995; Schneider, Einhäuser, & Horstmann, 2013). Adopting this view, we can interpret the present results on the allocation of attention and gaze during scene viewing as follows: the currently fixated location competes for attention with potential future locations. Once one of the competing peripheral locations outweighs the current location, a saccade toward the new location is executed. Hence less content at the current location and more content at peripheral locations both shorten fixation durations, as is observed in the present study.

The interaction between foveal and extrafoveal processing load provides clues about the nature of the parafoveal-on-foveal effect. According to the processing difficulty hypothesis, patches selected for fixation $n + 1$ may exert their influence on the duration of fixation n via parafoveal preprocessing, which is thought to be easier for unmodified than for isoluminant patches. Alternatively, unmodified patches in the periphery may be stronger "competitors" for attention and gaze than isoluminant patches. The fact that fixation durations in the "isoluminant to unmodified" condition were significantly shorter than fixation durations on completely unmodified images appears to support the latter view.

In the LATEST model of gaze control in scene viewing (Tatler et al., 2017), the decision to move the eyes is the result of a comparison between competing Stay and Go hypotheses: the relative benefit offered by moving the eyes to a new peripheral location or by staying at the currently fixated location is continuously evaluated. This implies that both foveal and extrafoveal information can influence fixation durations. Qualitatively, our results are well in line with these assumptions. Importantly, the LATEST model explicitly assumes an *inverse* relation between Stay and Go (Tatler et al., 2017). Some evidence for this model prediction comes from corpus-based analyses using a statistical control approach. When successor effects were found, they oftentimes had a sign opposite that for the corresponding immediacy effect in the linear mixed model (Nuthmann, 2017; Tatler et al., 2017). In the present Experiment 2, the immediacy and parafoveal-on-foveal effects were opposite in direction, lending further support to a central prediction of the LATEST model.

In the literature on eye-movement control in reading, parafoveal-on-foveal and successor effects are of theoretical importance as they allow for distinguishing between serial and parallel processing (Murray, Fischer, & Tatler, 2013). By comparison, the issue of parallel versus serial processing has received little empirical and theoretical investigation in scene perception (Nuthmann & Henderson, 2012). The spatial decision of where to fixate next likely involves some degree of parallel processing to identify and select candidates for fixation (see earlier text). Indeed, in the LATEST model saccades actually result from a race between *multiple* Stay-or-Go evaluations carried out *in parallel* across candidate locations in the visual field (Tatler et al., 2017). When investigating both selection in space and selection in time in Experiment 2, we found a parafoveal-on-foveal effect. Such a finding is naturally

compatible with the notion of parallel processing in scene viewing. Future research could investigate whether the effect extends beyond location $n + 1$, and whether the present pattern of results generalizes across different viewing tasks.

On a more general level, the competition between the currently fixated location and potential next locations can be understood as an instance of the exploitation-exploration dilemma (Cohen, McClure, & Yu, 2007). In the context of scene viewing, this means that the need to process the current location further (exploit) has to be weighed against visiting alternative locations (explore), an idea that has recently been investigated experimentally with a gaze-contingent guided-viewing task (Ehinger, Kaufhold, & König, 2018).

The pattern of results for the checkerboard stimuli in Experiment 2 suggests that observers do not always extend their fixation durations in conditions of increased processing difficulty, which is consistent with effects of spatial-frequency filtering on fixation durations during real-world scene perception and search (Cajar, Engbert, et al., 2016; Cajar, Schneeweiß, et al., 2016; Laubrock et al., 2013). To account for the counterintuitive finding of increased fixation durations for foveal high-pass and peripheral low-pass filtering, Laubrock et al. (2013) introduced a variant of the CRISP model (Nuthmann et al., 2010) in which foveal inhibition and peripheral disinhibition of the random saccade timer dynamically interact. As another critical feature, the model assumes that foveal and peripheral information processing evolve in parallel and independently (see also Ludwig, Davies, & Eckstein, 2014).

From the present data alone, it remains open whether it is the reduced low-level content or the affected informativeness of the patches that controls fixation duration. In the realm of predicting fixation probability, it has been argued that "informativeness" (Antes, 1974; Mackworth & Morandi, 1967), interestingness (Masciocchi, Mihalas, Parkhurst, & Niebur, 2009), "relevancy" (Onat, Açık, Schumann, & König, 2014), human-defined salience (Koehler, Guo, Zhang, & Eckstein, 2014) or "meaning" (Henderson, Hayes, Peacock, & Rehrig, 2019; see also Tatler et al., 2017) as judged by human observers guide gaze more effectively than image features. Similarly, human defined objects override low-level features (Stoll, Thrun, Nuthmann, & Einhäuser, 2015). Given the recent success of image-computable models that explicitly or implicitly incorporate object content in predicting fixation probability (Huang, Shen, Boix, & Zhao, 2015; Kümmerer, Wallis, & Bethge, 2016) or scanpaths (Adeli & Zelinsky, 2018), however, the differentiation between high-level or semantic content on the one hand and low-level features on the other hand might eventually become void, and should be replaced by a notion of image-computability. Extending such models, which currently have a strong focus on fixation probability, to predict fixation durations will be an interesting issue for future research. Our current data clearly highlight that the question as to where to fixate next should not be decoupled from the question as to how long to fixate here. Hence any successful model of gaze guidance should be able to predict the "where" *and* the "when" of fixations.

Keywords: attention, gaze, eye movements, features, salience, scene viewing

Acknowledgments

The authors thank Monique Michl for support in collecting the data.

The publication of this article was funded by Chemnitz University of Technology.

Commercial relationships: none.
Corresponding author: Wolfgang Einhäuser.
Email: wolfgang.einhaeuser-treyer@physik.tu-chemnitz.de.
Address: Chemnitz University of Technology, Institute of Physics–Physics of Cognition, Chemnitz, Germany.

Footnotes

[1]In this context, the fixated word (n) is typically referred to as the foveal word, and the word(s) adjacent to the fixated word ($n + x$ with $x \geq 1$) as the parafoveal word(s). Most of the time, but not always, this usage is consistent with the physiologically based definitions of the foveal and parafoveal regions of the visual field (Hyönä, 2011).
[2]It has been suggested to reserve the term parafoveal-on-foveal effect for evidence obtained with an experimental approach. When conducting corpus studies using a statistical control approach, the term successor effect should be used instead (Angele, Schotter, Slattery, Tenenbaum, Bicknell, & Rayner, 2015; Kliegl et al., 2006, for discussion). This is why the label successor effect was used in the corpus study by Nuthmann (2017), whereas we use the term parafoveal-on-foveal effect for our Experiment 2.

References

Adeli, H., & Zelinsky, G. (2018). Deep-BCN: Deep networks meet biased competition to create a brain-inspired model of attention control. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2045–2055, https://doi.org/10.1109/CVPRW.2018.00259.

Angele, B., Schotter, E. R., Slattery, T. J., Tenenbaum, T. L., Bicknell, K., & Rayner, K. (2015).

Do successor effects in reading reflect lexical parafoveal processing? Evidence from corpus-based and experimental eye movement data. *Journal of Memory and Language, 79-80*, 76–96, https://doi.org/10.1016/j.jml.2014.11.003.

Antes, J. R. (1974). The time course of picture viewing. *Journal of Experimental Psychology, 103*(1), 62–70, https://doi.org/10.1037/h0036799.

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

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

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433–436, https://doi.org/10.1163/156856897X00357.

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

Cajar, A., Engbert, R., & Laubrock, J. (2016). Spatial frequency processing in the central and peripheral visual field during scene viewing. *Vision Research, 127*, 186–197, https://doi.org/10.1016/j.visres.2016.05.008.

Cajar, A., Schneeweiß, P., Engbert, R., & Laubrock, J. (2016). Coupling of attention and saccades when viewing scenes with central and peripheral degradation. *Journal of Vision, 16*(2):8, 1–19, https://doi.org/10.1167/16.2.8.

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B-Biological Sciences, 362*(1481), 933–942, https://doi.org/10.1098/rstb.2007.2098.

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers, 34*(4), 613–617, https://doi.org/10.3758/BF03195489.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*(1), 193–222, https://doi.org/10.1146/annurev.neuro.18.1.193.

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

Drieghe, D. (2011). Parafoveal-on-foveal effects on eye movements during reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 839–856). Oxford: Oxford University Press, https://doi.org/10.1093/oxfordhb/9780199539789.013.0046.

Einhäuser, W., & König, P. (2003). Does luminance-contrast contribute to a saliency map for overt visual attention? *European Journal of Neuroscience, 17*(5), 1089–1097, https://doi.org/10.1046/j.1460-9568.2003.02508.x.

Einhäuser, W., & Nuthmann, A. (2016). Salient in space, salient in time: Fixation probability predicts fixation duration during natural scene viewing. *Journal of Vision, 16*(11):13, 1–17, https://doi.org/10.1167/16.11.13.

Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision, 8*(14):18, 1–26, https://doi.org/10.1167/8.14.18.

Ehinger, B. V., Kaufhold, L., & König, P. (2018). Probing the temporal dynamics of the exploration–exploitation dilemma of eye movements. *Journal of Vision, 18*(3):6, 1–24, https://doi.org/10.1167/18.3.6.

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

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

Harding, G., & Bloj, M. (2010). Real and predicted influence of image manipulations on eye movements during scene recognition. *Journal of Vision, 10*(2):8, 1–17, https://doi.org/10.1167/10.2.8.

Henderson, J. M., & Ferreira, F. (1990). Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(3), 417–429, https://doi.org/10.1037/0278-7393.16.3.417.

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

Henderson, J. M., Nuthmann, A, & Luke, S. G. (2013). Eye movement control during scene viewing: Immediate effects of scene luminance on fixation durations. *Journal of Experimental Psychology: Human Perception and Performance, 39*(2), 318–322, https://doi.org/10.1037/a0031224.

Ho-Phuoc, T., Guyader, N., Landragin, F., & Guérin-Dugué, A. (2012). When viewing natural scenes, do abnormal colors impact on spatial or temporal parameters of eye movements? *Journal of Vision, 12*(2):4, 1–13, https://doi.org/10.1167/12.2.4.

Huang, X., Shen, C., Boix, X., & Zhao, Q. (2015). Salicon: Reducing the semantic gap in saliency prediction by adapting deep neural networks. *Proceedings of the IEEE International Conference on Computer Vision*, 262–270, https://doi.org/10.1109/ICCV.2015.38.

Hyönä, J. (2011). Foveal and parafoveal processing during reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 819–838). Oxford: Oxford University Press, https://doi.org/10.1093/oxfordhb/9780199539789.013.0045.

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics, 40*(6), 431–439, https://doi.org/10.3758/BF03208203.

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

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in psychtoolbox-3. *Perception, 36*(1), 14–14.

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General, 135*(1), 12–35, https://doi.org/10.1037/0096-3445.135.1.12.

Koehler, K., Guo, F., Zhang, S., & Eckstein, M. P. (2014). What do saliency models predict? *Journal of Vision, 14*(3):14, 1–27, https://doi.org/10.1167/14.3.14.

Kümmerer, M., Wallis, T. S., & Bethge, M. (2016). DeepGaze II: Reading fixations from deep features trained on object recognition. Preprint at https://arxiv.org/abs/1610.01563.

Laubrock, J., Cajar, A., & Engbert, R. (2013). Control of fixation duration during scene viewing by interaction of foveal and peripheral processing. *Journal of Vision, 13*(12):11, 1–20, https://doi.org/10.1167/13.12.11.

Loftus, G. R. (1985). Picture perception: Effects of luminance on available information and information-extraction rate. *Journal of Experimental Psychology: General, 114*(3), 342–356, https://doi.org/10.1037/0096-3445.114.3.342.

Loftus, G. R., Kaufman, L., Nishimoto, T., & Ruthruff, E. (1992). Effects of visual degradation on eye-fixation durations, perceptual processing, and long-term visual memory. In K. Rayner (Ed.), *Eye movements and visual cognition: scene perception and reading* (pp. 203–226). New York: Springer, https://doi.org/10.1007/978-1-4612-2852-3_12.

Ludwig, C. J. H., Davies, J. R., & Eckstein, M. P. (2014). Foveal analysis and peripheral selection during active visual sampling. *Proceedings of the National Academy of Sciences of the United States of America, 111*(2), E291–E299, https://doi.org/10.1073/pnas.1313553111.

Mackworth, N. H., & Morandi, A. J. (1967). The gaze selects informative details within pictures. *Perception & Psychophysics, 2*(11), 547–552, https://doi.org/10.3758/BF03210264.

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1995). Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-D images. *Spatial Vision, 9*(3), 363–386, https://doi.org/10.1163/156856895X00052.

Masciocchi, C. M., Mihalas, S., Parkhurst, D., & Niebur, E. (2009). Everyone knows what is interesting: Salient locations which should be fixated. *Journal of Vision, 9*(11):25, 1–22, https://doi.org/10.1167/9.11.25.

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics, 17*(6), 578–586, https://doi.org/10.3758/BF03203972.

Murray, W. S., Fischer, M. H., & Tatler, B. W. (2013). Serial and parallel processes in eye movement control: Current controversies and future directions. *The Quarterly Journal of Experimental Psychology, 66*(3), 417–428, https://doi.org/10.1080/17470218.2012.759979.

Nuthmann, A. (2013). On the visual span during object search in real-world scenes. *Visual Cognition, 21*(7), 803–837, https://doi.org/10.1080/13506285.2013.832449.

Nuthmann, A. (2014). How do the regions of the visual field contribute to object search in real-world scenes? Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 40*(1), 342–360, https://doi.org/10.1037/a0033854.

Nuthmann, A. (2017). Fixation durations in scene viewing: Modeling the effects of local image features, oculomotor parameters, and task. *Psychonomic Bulletin & Review, 24*(2), 370–392, https://doi.org/10.3758/s13423-016-1124-4.

Nuthmann, A., & Einhäuser, W. (2015). A new approach to modeling the influence of image features on fixation selection in scenes. *Annals of*

the New York Academy of Sciences, 1339, 82–96, https://doi.org/10.1111/nyas.12705.

Nuthmann, A., & Henderson, J. M. (2012). Using CRISP to model global characteristics of fixation durations in scene viewing and reading with a common mechanism. *Visual Cognition, 20*(4-5), 457–494, https://doi.org/10.1080/13506285.2012.670142.

Nuthmann, A., & Malcolm, G. L. (2016). Eye guidance during real-world scene search: The role color plays in central and peripheral vision. *Journal of Vision, 16*(2):3, 1–16, https://doi.org/10.1167/16.2.3.

Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review, 117*(2), 382–405, https://doi.org/10.1037/a0018924.

Nyström, M., & Holmqvist, K. (2008). Semantic override of low-level features in image viewing–Both initially and overall. *Journal of Eye Movement Research, 2*(2):2, 1–11, https://doi.org/10.16910/jemr.2.2.2.

Onat, S., Açık, A., Schumann, F., & König, P. (2014). The contributions of image content and behavioral relevancy to overt attention. *PLoS One, 9*(4), e93254, https://doi.org/10.1371/journal.pone.0093254.

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin, 85*(3), 618–660, https://doi.org/10.1037/0033-2909.85.3.618.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422, https://doi.org/10.1037//0033-2909.124.3.372.

Rayner, K. (2009). The 35th Sir Frederick Bartlett lecture: Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62*(8), 1457–1506, https://doi.org/10.1080/17470210902816461.

Rayner, K., & Bertera, J. H. (1979). Reading without a fovea. *Science, 206*(4417), 468–469, https://doi.org/10.1126/science.504987.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26*(4), 445–476, https://doi.org/10.1017/S0140525X03000104.

Reichle, E. D., & Reingold, E. M. (2013). Neurophysiological constraints on the eye-mind link. *Frontiers in Human Neuroscience, 7*, 361, https://doi.org/10.3389/fnhum.2013.00361.

Saez de Urabain, I. R., Nuthmann, A., Johnson, M. H., & Smith, T. J. (2017). Disentangling the mechanisms underlying infant fixation durations in scene perception: A computational account. *Vision Research, 134*, 43–59, https://doi.org/10.1016/j.visres.2016.10.015.

Schad, D. J., & Engbert, R. (2012). The zoom lens of attention: Simulating shuffled versus normal text reading using the SWIFT model. *Visual Cognition, 20*(4-5), 391–421, https://doi.org/10.1080/13506285.2012.670143.

Schneider, W. X., Einhäuser, W., & Horstmann, G. (2013). Attentional selection in visual perception, memory and action: A quest for cross-domain integration. *Philosophical Transactions of the Royal Society B, 368*(1628), 20130053, https://doi.org/10.1098/rstb.2013.0053.

Shore, S., Tillman, L., & Schmidt-Wulffen, S. (2004). *Stephen shore: Uncommon places: The complete works*. New York: Aperture.

Tatler, B. W., Brockmole, J. R., & Carpenter, R. H. S. (2017). LATEST: A model of saccadic decisions in space and time. *Psychological Review, 124*(3), 267–300, https://doi.org/10.1037/rev0000054.

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision, 11*(5):5, 1–23, https://doi.org/10.1167/11.5.5.

von Wartburg, R., Ouerhani, N., Pflugshaupt, T., Nyffeler, T., Wurtz, P., Hügli, H., … Müri, R. M. (2005). The influence of colour on oculomotor behaviour during image perception. *Neuroreport, 16*(14), 1557–1560, https://doi.org/10.1097/01.wnr.0000180146.84020.c4.

Walshe, R. C., & Nuthmann, A. (2015). Mechanisms of saccadic decision making while encoding naturalistic scenes. *Journal of Vision, 15*(5):21, 1–19, https://doi.org/10.1167/15.5.21.

Williams, C. C., & Castelhano, M. S. (2019). The changing landscape: High-level influences on eye movement guidance in scenes. *Vision, 3*(3):133, 1–20, https://doi.org/10.3390/vision3030033.